

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Thomas, Winston J.
Drayna, Dennis T.
Feder, John N.
Gnirke, Andreas
Ruddy, David
Tsuchihashi, Zenta
Wolff, Roger K.

(ii) TITLE OF INVENTION: Hereditary Hemochromatosis Gene

(iii) NUMBER OF SEQUENCES: 44

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Townsend and Townsend and Crew LLP
(B) STREET: Two Embarcadero Center, Eighth Floor
(C) CITY: San Francisco
(D) STATE: California
(E) COUNTRY: USA
(F) ZIP: 94111-3834

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US 08/652,265
(B) FILING DATE: 23-MAY-1996
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Smith, William M.
(B) REGISTRATION NUMBER: 30,223
(C) REFERENCE/DOCKET NUMBER: 17957-000500

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (415) 576-0200
(B) TELEFAX: (415) 576-0300

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10825 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: join(361..436, 3762..4025, 4235..4510, 5606..5881,
6040..6153, 7107..7147)
(D) OTHER INFORMATION: /product= "Hereditary Hemochromatosis"

(HH) protein"
 /note= "Normal or wild-type (unaffected)
 Hereditary Hemochromatosis (HH) gene
 allele"

(ix) FEATURE:

(A) NAME/KEY: -
 (B) LOCATION: 140..7319
 (D) OTHER INFORMATION: /note= "start and stop positions for
 normal or wild-type (unaffected) allele
 cDNA (SEQ ID NO:9)"

(ix) FEATURE:

(A) NAME/KEY: -
 (B) LOCATION: 3852..3891
 (D) OTHER INFORMATION: /note= "start and stop positions for
 normal or wild-type (unaffected) genomic
 sequence surrounding variant for 24d2(C)
 allele (SEQ ID NO:41)"

(ix) FEATURE:

(A) NAME/KEY: -
 (B) LOCATION: 5507..6023
 (D) OTHER INFORMATION: /note= "start and stop positions for
 normal or wild-type (unaffected) genomic
 sequence surrounding variant for 24d1(G)
 allele (SEQ ID NO:20)"

(ix) FEATURE:

(A) NAME/KEY: allele
 (B) LOCATION: replace(3872, "c")
 (D) OTHER INFORMATION: /phenotype= "normal or wild-type
 (unaffected)"
 /label= 24d2

(ix) FEATURE:

(A) NAME/KEY: allele
 (B) LOCATION: replace(3878, "a")
 (D) OTHER INFORMATION: /phenotype= "normal or wild-type
 (unaffected)"
 /label= 24d7

(ix) FEATURE:

(A) NAME/KEY: allele
 (B) LOCATION: replace(5834, "g")
 (D) OTHER INFORMATION: /phenotype= "normal or wild-type
 (unaffected)"
 /label= 24d1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TCTAAGGTTG AGATAAAATT TTTAAATGTA TGATTGAATT TTGAAAATCA TAAATATTTA	60
AATATCTAAA GTTCAGATCA GAACATTGCG AAGCTACTTT CCCCCAATCAA CAACACCCCT	120
TCAGGATTCA AAAACCAAGG GGGACACTGG ATCACCTAGT GTTTCACAAG CAGGTACCTT	180
CTGCTGTAGG AGAGAGAGAA CTAAAGTTCT GAAAGACCTG TTGCTTTCA CCAGGAAGTT	240
TTACTGGGCA TCTCCTGAGC CTAGGCAATA GCTGTAGGGT GACTTCTGGA GCCATCCCCG	300

TTTCCCCGCC	CCCCAAAAGA	AGCGGAGATT	TAACGGGGAC	GTGCGGCCAG	AGCTGGGGAA	360
ATG GGC CCG CGA GCC AGG CCG GCG CTT CTC CTC CTG ATG CTT TTG CAG						408
Met Gly Pro Arg Ala Arg Pro Ala Leu Leu Leu Leu Met Leu Leu Gln						
1	5	10	15			
ACC GCG GTC CTG CAG GGG CGC TTG CTG C	GTGAGTCCGA	GGGCTGCGGG				456
Thr Ala Val Leu Gln Gly Arg Leu Leu						
20	25					
CGAACTAGGG GCGCGGCCGG GGTGGAAAAA TCGAAACTAG	CTTTTCTTT	GCGCTTGGGA				516
GTGGCTAAC TTTGGAGGAC CTGCTCAACC CTATCCGCAA	GCCCCCTCTCC	CTACTTTCTG				576
CGTCCAGACC CGGTGAGGGA GTGCCTACCA CTGAAC TGCA	GATAGGGGTC	CCTCGCCCCA				636
GGACCTGCCCG CCTCCCCCGG CTGTCCCGC TCTGCGGAGT	GACTTTGGA	ACCGCCCAC				696
CCCTTCCCCC AACTAGAATG CTTTAAATA AATCTCGTAG	TTCCCTCACTT	GAGCTGAGCT				756
AAGCCTGGGG CTCCTTGAAC CTGGAAC TCGAGTCG	GGTTTATTTC	CAATGTCAGC	TGTGCAGTT			816
TTTCCCCAGT CATCTCCAAA CAGGAAGTTC TTCCCTGAGT	GCTTGCCGAG	AAGGCTGAGC				876
AAACCCACAG CAGGATCCGC ACGGGGTTTC CACCTCAGAA	CGAATGCGTT	GGCGGGTGGG				936
GGCGCGAAAG AGTGGCGTTG GGGATCTGAA TTCTTCACCA	TTCCACCCAC	TTTGGTGAG				996
ACCTGGGTG GAGGTCTCTA GGGTGGGAGG CTCCTGAGAG	AGGCCTACCT	CGGGCCTTTC				1056
CCCACTCTG GCAATTGTTTC TTTGCCTGG AAAATTAAGT	ATATGTTAGT	TTGAACGTT				1116
TGAACTGAAC AATTCTCTTT TCGGCTAGGC TTTATTGATT	TGCAATGTGC	TGTGTAATT				1176
AGAGGCCTCT CTACAAAGTA CTGATAATGA ACATGTAAGC	AATGCACTCA	CTTCTAAGTT				1236
ACATTCAATAT CTGATCTTAT TTGATTTCA CTAGGCATAG	GGAGGTAGGA	GCTAATAATA				1296
CGTTTATTTT ACTAGAAGTT AACTGGAATT CAGATTATAT	AACTCTTTTC	AGGTTACAAA				1356
GAACATAAAAT AATCTGGTTT TCTGATGTTA TTTCAAGTAC	TACAGCTGCT	TCTAATCTTA				1416
GTTGACAGTG ATTTGCCCT GTAGTGTAGC ACAGTGTCT	GTGGGTACACA	CGCCGGCCTC				1476
AGCACAGCAC TTTGAGTTTT GGTACTACGT GTATCCACAT	TTTACACATG	ACAAGAATGA				1536
GGCATGGCAC GGCCTGCTTC CTGGCAAATT TATTCAATGG	TACACTGGGC	TTTGGTGGCA				1596
GAGCTCATGT CTCCACTTCA TAGCTATGAT TCTTAAACAT	CACACTGCAT	TAGAGGTTGA				1656
ATAATAAAAT TTCAATGTTGA GCAGAAATAT TCATTGTTA	CAAGTGTAAA	TGAGTCCCAG				1716
CCATGTGTTG CACTGTTCAA GCCCCAAGGG AGAGAGCAGG	GAAACAAGTC	TTTACCCCTT				1776
GATATTTGC ATTCTAGTGG GAGAGATGAC AATAAGCAA	TGAGCAGAAA	GATATACAAC				1836
ATCAGGAAAT CATGGGTGTT GTGAGAAGCA GAGAAGTCAG	GGCAAGTCAC	TCTGGGGCTG				1896
ACACTTGAGC AGAGACATGA AGGAAATAAG AATGATATTG	ACTGGGAGCA	GTATTTCCCA				1956
GGCAAACATGA GTGGGCCTGG CAAGTTGGAT TAAAAAGCGG	GTTTTCTCAG	CACTACTCAT				2016

GTGTGTGTGT	GTGGGGGGGG	GGGGCGGCGT	GGGGGTGGGA	AGGGGGACTA	CCATCTGCAT	2076
GTAGGATGTC	TAGCA GTATC	CTGT CCTCCC	TACTCACTAG	GTGCTAGGAG	CACTCCCCA	2136
GTCTTGACAA	CCAAAATGT	CTCTAAACTT	TGCCACATGT	CACCTAGTAG	ACAAACTCCT	2196
GGTTAAGAAG	CTCGGGTTGA	AAAAAATAAA	CAAGTAGTGC	TGGGGAGTAG	AGGCCAAGAA	2256
GTAGGTAATG	GGCTCAGAAG	AGGAGCCACA	AACAAGGTTG	TGCAGGCC	TGTAGGCTGT	2316
GGTGTGAATT	CTAGCCAAGG	AGTAACAGTG	ATCTGTACAC	GGCTTTAAA	AGATTGCTCT	2376
GGCTGCTATG	TGGAAAGCAG	AATGAAGGGA	GCAACAGTAA	AAGCAGGGAG	CCCAGCCAGG	2436
AAGCTGTTAC	ACAGTCCAGG	CAAGAGGTAG	TGGAGTGGC	TGGGTGGAA	CAGAAAAGGG	2496
AGTGACAAAC	CATTGTCTCC	TGAATATATT	CTGAAGGAAG	TTGCTGAAGG	ATTCTATGTT	2556
GTGTGAGAGA	AAGAGAAGAA	TTGGCTGGGT	GTAGTAGCTC	ATGCCAAGGA	GGAGGCCAAG	2616
GAGAGCAGAT	TCCTGAGCTC	AGGAGTTCAA	GACCAGCCTG	GGCAACACAG	CAAAACCCCT	2676
TCTCTACAAA	AAATACAAAAA	ATTAGCTGGG	TGTGGTGGCA	TGCACCTGTG	ATCCTAGCTA	2736
CTCAGGGAGGC	TGAGGTGGAG	GGTATTGCTT	GAGCCCAGGA	AGTTGAGGCT	GCAGTGAGCC	2796
ATGACTGTGC	CACTGTACTT	CAGCCTAGGT	GACAGAGCAA	GACCCTGTCT	CCCCTGACCC	2856
CCTGAAAAAG	AGAAGAGTTA	AAGTTGACTT	TGTTCTTTAT	TTTAATTTTA	TTGGCCTGAG	2916
CAGTGGGGTA	ATTGGCAATG	CCATTCTGA	GATGGTGAAG	GCAGAGGAAA	GAGCAGTTTG	2976
GGGTAAATCA	AGGATCTGCA	TTTGGGACAT	GTAAAGTTG	AGATTCCAGT	CAGGCTTCCA	3036
AGTGGTGAGG	CCACATAGGC	AGTTCA GTGT	AAGAATTCA G	GACCAAGGCT	GGGCACGGTG	3096
GCTCACTTCT	GTAATCCCAG	CACTTTGGTG	GCTGAGGCAG	GTAGATCATT	TGAGGTCAGG	3156
AGTTTGAGAC	AAGCTTGGCC	AACATGGTGA	AACCCATGT	CTACTAAAAA	TACAAAATT	3216
AGCCTGGTGT	GGTGGCGCAC	GCCTATAGTC	CCAGGTTTTC	AGGAGGCTTA	GGTAGGAGAA	3276
TCCCTTGAAC	CCAGGAGGTG	CAGGTTGCAG	TGAGCTGAGA	TTGTGCCACT	GCAC TCCAGC	3336
CTGGGTGATA	GAGTGAGACT	CTGTCTCAA	AAAAAAAAAA	AAAAAAAAAA	AAAAAAACTGA	3396
AGGAATTATT	CCTCAGGATT	TGGGTCTAAT	TTGCCCTGAG	CACCAACTCC	TGAGTTCAAC	3456
TACCATGGCT	AGACACACCT	TAACATTTTC	TAGAATCCAC	CAGCTTAGT	GGAGTCTGTC	3516
TAATCATGAG	TATTGGAATA	GGATCTGGGG	GCAGTGAGGG	GGTGGCAGCC	ACGTGTGGCA	3576
GAGAAAAGCA	CACAAGGAAA	GAGCACCCAG	GACTGTCATA	TGGAAGAAAG	ACAGGACTGC	3636
AACTCACCCCT	TCACAAAATG	AGGACCAGAC	ACAGCTGATG	GTATGAGTTG	ATGCAGGTGT	3696
GTGGAGCCTC	AACATCCTGC	TCCCCTCCTA	CTACACATGG	TTAAGGCCTG	TTGCTCTGTC	3756
TCCAG	GT TCA CAC TCT	CTG CAC TAC	CTC TTC ATG	GGT GCC TCA	GAG	3802
	Arg Ser His Ser	Leu His Tyr	Leu Phe Met	Gly Ala Ser	Glu	

CAG GAC CTT GGT CTT TCC TTG TTT GAA GCT TTG GGC TAC GTG GAT GAC Gln Asp Leu Gly Leu Ser Leu Phe Glu Ala Leu Gly Tyr Val Asp Asp 40 45 50 55	3850
CAG CTG TTC GTG TTC TAT GAT CAT GAG AGT CGC CGT GTG GAG CCC CGA Gln Leu Phe Val Phe Tyr Asp His Glu Ser Arg Arg Val Glu Pro Arg 60 65 70	3898
ACT CCA TGG GTT TCC AGT AGA ATT TCA AGC CAG ATG TGG CTG CAG CTG Thr Pro Trp Val Ser Ser Arg Ile Ser Ser Gln Met Trp Leu Gln Leu 75 80 85	3946
AGT CAG AGT CTG AAA GGG TGG GAT CAC ATG TTC ACT GTT GAC TTC TGG Ser Gln Ser Leu Lys Gly Trp Asp His Met Phe Thr Val Asp Phe Trp 90 95 100	3994
ACT ATT ATG GAA AAT CAC AAC CAC AGC AAG G GTATGTGGAG AGGGGGCCTC Thr Ile Met Glu Asn His Asn His Ser Lys 105 110	4045
ACCTTCCTGA GGTTGTCAGA GCTTTTCATC TTTTCATGCA TCTTGAAGGA AACAGCTGGA	4105
AGTCTGAGGT CTTGTGGAG CAGGGAAAGAG GGAAGGAATT TGCTTCCTGA GATCATTG	4165
TCCTTGGGGA TGGTGGAAAT AGGGACCTAT TCCTTGGTT GCAGTTAACAGGCTGGGA	4225
TTTTTCCAG AG TCC CAC ACC CTG CAG GTC ATC CTG GGC TGT GAA ATG Glu Ser His Thr Leu Gln Val Ile Leu Gly Cys Glu Met 115 120 125	4272
CAA GAA GAC AAC AGT ACC GAG GGC TAC TGG AAG TAC GGG TAT GAT GGG Gln Glu Asp Asn Ser Thr Glu Gly Tyr Trp Lys Tyr Gly Tyr Asp Gly 130 135 140	4320
CAG GAC CAC CTT GAA TTC TGC CCT GAC ACA CTG GAT TGG AGA GCA GCA Gln Asp His Leu Glu Phe Cys Pro Asp Thr Leu Asp Trp Arg Ala Ala 145 150 155	4368
GAA CCC AGG GCC TGG CCC ACC AAG CTG GAG TGG GAA AGG CAC AAG ATT Glu Pro Arg Ala Trp Pro Thr Lys Leu Glu Trp Glu Arg His Lys Ile 160 165 170	4416
CGG GCC AGG CAG AAC AGG GCC TAC CTG GAG AGG GAC TGC CCT GCA CAG Arg Ala Arg Gln Asn Arg Ala Tyr Leu Glu Arg Asp Cys Pro Ala Gln 175 180 185 190	4464
CTG CAG CAG TTG CTG GAG CTG GGG AGA GGT GTT TTG GAC CAA CAA G Leu Gln Gln Leu Leu Glu Leu Gly Arg Gly Val Leu Asp Gln Gln 195 200 205	4510
GTATGGTGGAA AACACACTTC TGCCCCTATA CTCTAGTGGC AGAGTGGAGG AGGTTGCAGG	4570
GCACGGAATC CCTGGTTGGA GTTTCAGAGG TGGCTGAGGC TGTGTGCCTC TCCAAATTCT	4630
GGGAAGGGAC TTTCTCAATC CTAGAGTCTC TACCTTATAA TTGAGATGTA TGAGACAGCC	4690
ACAAGTCATG GGTTTAATTCTTCTTCTCCA TGCATATGGC TCAAAGGGAA GTGTCTATGG	4750
CCCTTGCTTT TTATTTAACCA AATAATCTTT TGTATATTAA TACCTGTTAA AAATTCAGAA	4810
ATGTCAAGGC CGGGCACGGT GGCTCACCCCC TGTAATCCCA GCACTTGGG AGGCCGAGGC	4870

GGGTGGTCAC AAGGTCAGGA GTTTGAGACC AGCCTGACCA ACATGGTGAA ACCCGTCTCT	4930
AAAAAAATAC AAAAATTAGC TGGTCACAGT CATGCGCACC TGTAGTCCCA GCTAATTGGA	4990
AGGCTGAGGC AGGAGCATCG CTTGAACCTG GGAAGCGGAA GTTGCACTGA GCCAAGATCG	5050
CGCCACTGCA CTCCAGCCTA GGCAGCAGAG TGAGACTCCA TCTTAAAAAA AAAAAAAA	5110
AAAAAAAGAG AATTCAGAGA TCTCAGCTAT CATATGAATA CCAGGACAAA ATATCAAGTG	5170
AGGCCACTTA TCAGAGTAGA AGAACATCTT AGGTTAAAAG TTTCTTTCAT AGAACATAGC	5230
ATAAATCACT GAAGCTACCT ATCTTACAAG TCCGCTTCTT ATAACAATGC CTCCTAGGTT	5290
GACCCAGGTG AAACGTACCA TCTGTATTCA ATCATTTCA ATGCACATAA AGGGCAATTT	5350
TATCTATCAG AACAAAGAAC ATGGGTAACA GATATGTATA TTTACATGTG AGGAGAACAA	5410
GCTGATCTGA CTGCTCTCCA AGTGACACTG TGTTAGAGTC CAATCTTAGG ACACAAAATG	5470
GTGTCTCTCC TGAGCTTGT TTTTTCTGA AAAGGGTATT TCCTTCCTCC AACCTATAGA	5530
AGGAAGTGAA AGTTCCAGTC TTCCGGCAA GGGTAAACAG ATCCCCCTCTC CTCATCCTTC	5590
CTCTTCCCTG TCAAG TG CCT CCT TTG GTG AAG GTG ACA CAT CAT GTG ACC Val Pro Pro Leu Val Lys Val Thr His His Val Thr	5640
210 215	
TCT TCA GTG ACC ACT CTA CGG TGT CGG GCC TTG AAC TAC TAC CCC CAG Ser Ser Val Thr Leu Arg Cys Arg Ala Leu Asn Tyr Tyr Pro Gln	5688
220 225 230	
AAC ATC ACC ATG AAG TGG CTG AAG GAT AAG CAG CCA ATG GAT GCC AAG Asn Ile Thr Met Lys Trp Leu Lys Asp Lys Gln Pro Met Asp Ala Lys	5736
235 240 245	
GAG TTC GAA CCT AAA GAC GTA TTG CCC AAT GGG GAT GGG ACC TAC CAG Glu Phe Glu Pro Lys Asp Val Leu Pro Asn Gly Asp Gly Thr Tyr Gln	5784
250 255 260 265	
GGC TGG ATA ACC TTG GCT GTA CCC CCT GGG GAA GAG CAG AGA TAT ACG Gly Trp Ile Thr Leu Ala Val Pro Pro Gly Glu Glu Gln Arg Tyr Thr	5832
270 275 280	
TGC CAG GTG GAG CAC CCA GGC CTG GAT CAG CCC CTC ATT GTG ATC TGG G Cys Gln Val Glu His Pro Gly Leu Asp Gln Pro Leu Ile Val Ile Trp	5881
285 290 295	
GTATGTGACT GATGAGAGCC AGGAGCTGAG AAAATCTATT GGGGGTTGAG AGGAGTGCCT	5941
GAGGAGGTAA TTATGGCAGT GAGATGAGGA TCTGCTCTTT GTTAGGGGGT GGGCTGAGGG	6001
TGGCAATCAA AGGCTTTAAC TTGCTTTTC TGTTTAG AG CCC TCA CCG TCT Glu Pro Ser Pro Ser	6053
300	
GGC ACC CTA GTC ATT GGA GTC ATC AGT GGA ATT GCT GTT TTT GTC GTC Gly Thr Leu Val Ile Gly Val Ile Ser Gly Ile Ala Val Phe Val Val	6101
305 310 315	
ATC TTG TTC ATT GGA ATT TTG TTC ATA ATA TTA AGG AAG AGG CAG GGT Ile Leu Phe Ile Gly Ile Leu Phe Ile Leu Arg Lys Arg Gln Gly	6149

320	325	330	
TCA A GTGAGTAGGA ACAAGGGGGA AGTCTCTTAG TACCTCTGCC CCAGGGCACA			6203
Ser			
335			
GTGGGAAGAG GGGCAGAGGG GATCTGGCAT CCATGGGAAG CATTTCCTTC ATTATATATTC			6263
TTTGGGACA CCAGCAGCTC CCTGGGAGAC AGAAAATAAT GGTTCTCCCC AGAATGAAAG			6323
TCTCTAATTC AACAAACATC TTCAGAGCAC CTACTATTT GCAAGAGCTG TTTAAGGTAG			6383
TACAGGGCCT TTGAGGTTGA GAAGTCACTG TGGCTATTCT CAGAACCCAA ATCTGGTAGG			6443
GAATGAAATT GATAGCAAGT AAATGTAGTT AAAGAAGACC CCATGAGGTC CTAAAGCAGG			6503
CAGGAAGCAA ATGCTTAGGG TGTCAAAGGA AAGAATGATC ACATTCAAGCT GGGGATCAAG			6563
ATAGCCTTCT GGATCTTGAA GGAGAAGCTG GATTCCATTA GGTGAGGTTG AAGATGATGG			6623
GAGGTCTACA CAGACGGAGC AACCATGCCA AGTAGGAGAG TATAAGGCAT ACTGGGAGAT			6683
TAGAAATAAT TACTGTACCT TAACCTGAG TTTGCGTAGC TATCACTCAC CAATTATGCA			6743
TTTCTACCCC CTGAACATCT GTGGTGTAGG GAAAAGAGAA TCAGAAAGAA GCCAGCTCAT			6803
ACAGAGTCCA AGGGTCTTTT GGGATATTGG GTTATGATCA CTGGGGTGTC ATTGAAGGAT			6863
CCTAAGAAAG GAGGACCACG ATCTCCCTTA TATGGTGAAT GTGTTGTTAA GAAGTTAGAT			6923
GAGAGGTGAG GAGACCAGTT AGAAAGCAA TAAGCATTTC CAGATGAGAG ATAATGGTTC			6983
TTGAAATCCA ATAGTGCCTA GGTCTAAATT GAGATGGGTG AATGAGGAAA ATAAGGAAGA			7043
GAGAAGAGGC AAGATGGTGC CTAGGTTGT GATGCCTCTT TCCTGGGTCT CTTGTCTCCA			7103
CAG GA GGA GCC ATG GGG CAC TAC GTC TTA GCT GAA CGT GAG			7144
Arg Gly Ala Met Gly His Tyr Val Leu Ala Glu Arg Glu			
340	345		
TGACACGCAG CCTGCAGACT CACTGTGGGA AGGAGACAAA ACTAGAGACT CAAAGAGGGA			7204
GTGCATTTAT GAGCTCTCA TGTTTCAGGA GAGAGTTGAA CCTAAACATA GAAATTGCCT			7264
GACGAACCTCC TTGATTTAG CCTTCTCTGT TCATTTCTC AAAAAGATT CCCATTTAG			7324
GTTCCTGAGT TCCTGCATGC CGGTGATCCC TAGCTGTGAC CTCTCCCTG GAACTGTCTC			7384
TCATGAACCT CAAGCTGCAT CTAGAGGCTT CCTTCATTTC CTCCGTACCC TCAGAGACAT			7444
ACACCTATGT CATTTCATTT CCTATTTTG GAAGAGGACT CCTTAAATTT GGGGGACTTA			7504
CATGATTCAAT TTTAACATCT GAGAAAAGCT TTGAACCTG GGACGTGGCT AGTCATAACC			7564
TTACCCAGATT TTTACACATG TATCTATGCA TTTCTGGAC CCGTTCAACT TTTCCCTTGA			7624
ATCCTCTCTC TGTGTTACCC AGTAACTCAT CTGTCACCAA GCCTTGGGGA TTCTTCCATC			7684
TGATTGTGAT GTGAGTTGCA CAGCTATGAA GGCTGTACAC TGCACGAATG GAAGAGGCAC			7744
CTGTCCCAGA AAAAGCATCA TGGCTATCTG TGGTAGTAT GATGGGTGTT TTTAGCAGGT			7804

AGGAGGCAAA TATCTTGAAA GGGGTTGTGA AGAGGTGTTT TTTCTAATTG GCATGAAGGT	7864
GTCATACAGA TTTGCAAAGT TTAATGGTGC CTTCATTTGG GATGCTACTC TAGTATTCCA	7984
GACCTGAAGA ATCACAATAA TTTTCTACCT GGTCTCTCCT TGTTCTGATA ATGAAAATTA	7984
TGATAAGGAT GATAAAAGCA CTTACTTCGT GTCCGACTCT TCTGAGCACC TACTTACATG	8044
CATTACTGCA TGCACCTCTT ACAATAATTC TATGAGATAG GTACTATTAT CCCCATTTCT	8104
TTTTTAAATG AAGAAAGTGA AGTAGGCCGG GCACGGTGGC TCACGCCTGT AATCCCAGCA	8164
CTTTGGGAGG CCAAAGCGGG TGGATCACGA GGTCAAGGAGA TCGAGACCAT CCTGGCTAAC	8224
ATGGTGAAAC CCCATCTCTA ATAAAAAATAC AAAAAAATTAG CTGGGCGTGG TGGCAGACGC	8284
CTGTAGTCCC AGCTACTCGG AAGGCTGAGG CAGGAGAATG GCATGAACCC AGGAGGCAGA	8344
GCTTGCAGTG AGCCGAGTTT GCGCCACTGC ACTCCAGCCT AGGTGACAGA GTGAGACTCC	8404
ATCTCAAAAA AATAAAAATA AAAATAAAA AATGAAAAAA AAAAGAAAGT GAAGTATAGA	8464
GTATCTCATA GTTTGTCAGT GATAGAAACA GGTTTCAAAC TCAGTCAATC TGACCGTTG	8524
ATACATCTCA GACACCACTA CATTCACTAG TTTAGATGCC TAGAATAAAT AGAGAAGGAA	8584
GGAGATGGCT CTTCTCTTGT CTCATTGTGT TTCTTCTGAG TGAGCTTGAA TCACATGAAG	8644
GGGAACAGCA GAAAACAACC AACTGATCCT CAGCTGTCA GTTTCCTTTA AAAGTCCCTG	8704
AAGGAAGGTC CTGGAATGTG ACTCCCTTGC TCCTCTGTTG CTCTCTTGG CATTCAATTTC	8764
TTTGGACCCCT ACGCAAGGAC TGTAATTGGT GGGGACAGCT AGTGGCCCTG CTGGGTTCA	8824
CACACGGTGT CCTCCCTAGG CCAGTGCCTC TGGAGTCAGA ACTCTGGTGG TATTTCCCTC	8884
AATGAAGTGG AGTAAGCTCT CTCATTTGA GATGGTATAA TGGAAAGCCAC CAAGTGGCTT	8944
AGAGGATGCC CAGGTCCCTTC CATGGAGCCA CTGGGGTTCC GGTGCACATT AAAAAAAA	9004
TCTAACCAAGG ACATTCAGGA ATTGCTAGAT TCTGGAAAT CAGTTCACCA TGTTCAAAAG	9064
AGTCTTTTTT TTTTTTTGGA GACTCTATTG CCCAGGCTGG AGTGAATGG CATGATCTCG	9124
GCTCACTGTA ACCTCTGCCT CCCAGGTTCA AGCGATTCTC CTGTCTCAGC CTCCCAAGTA	9184
GCTGGGATTA CAGGCAGTCA CCACCATGCC CGGCTAATTT TTGTATTTTT AGTAGAGACA	9244
GGGTTTCACC ATGTTGGCCA GGCTGGTCTC GAACTCTCCT GACCTCGTGA TCCGCCTGCC	9304
TCGGCCTCCC AAAGTGTGA GATTACAGGT GTGAGCCACC CTGCCAGCC GTCAAAAGAG	9364
TCTTAATATA TATATCCAGA TGGCATGTGT TTACTTTATG TTACTACATG CACTTGGCTG	9424
CATAAAATGTG GTACAAGCAT TCTGTCTTGA AGGGCAGGTG CTTCAAGGATA CCATATACAG	9484
CTCAGAAGTT TCTTCTTTAG GCATTAATT TTAGCAAAGA TATCTCATCT CTTCTTTAA	9544
ACCATTTCT TTTTTGTGG TTAGAAAAGT TATGTAGAAA AAAGTAAATG TGATTTACGC	9604
TCATTGTAGA AAAGCTATAA AATGAATACA ATTAAGCTG TTATTTAATT AGCCAGTGAA	9664

AAACTATTAA CAACTTGTCT ATTACCTGTT AGTATTATTG TTGCATTAAA AATGCATATA	9724
CTTTAATAAA TGTATATTGT ATTGTATACT GCATGATTT ATTGAAGTTC TTGTTCATCT	9784
TGTGTATATA CTTAATCGCT TTGTCATTTT GGAGACATTT ATTTGCTTC TAATTTCTTT	9844
ACATTTGTC TTACGGAATA TTTTCATTCA ACTGTGGTAG CGAATTAAT CGTGTTC	9904
CACTCTAGGG ACATTGTCGT CTAAGTTGTA AGACATTGGT TATTTTACCA GCAAACCATT	9964
CTGAAAGCAT ATGACAAATT ATTTCTCTCT TAATATCTTA CTATACTGAA AGCAGACTGC	10024
TATAAGGCTT CACTTACTCT TCTACCTCAT AAGGAATATG TTACAATTAA TTTATTAGGT	10084
AAGCATTGTT TTTATATTGG TTTTATTCA CCTGGGCTGA GATTCAAGA AACACCCAG	10144
TCTTCACAGT AACACATTTC ACTAACACAT TTACTAAACA TCAGCAACTG TGGCCTGTTA	10204
ATTTTTTAA TAGAAATTAA AAGTCCTCAT TTTCTTCGG TGTTTTTAA GCTTAATTAA	10264
TCTGGCTTTA TTCATAAATT CTTAAGGTCA ACTACATTG AAAAATCAAA GACCTGCATT	10324
TTAAATTCTT ATTCACCTCT GGCAAAACCA TTCACAAACC ATGGTAGTAA AGAGAAGGGT	10384
GACACCTGGT GGCCATAGGT AAATGTACCA CGGTGGTCCG GTGACCAGAG ATGCAGCGCT	10444
GAGGGTTTTC CTGAAGGTAA AGGAATAAAG AATGGGTGGA GGGCGTGCA CTGAAATCA	10504
CTTGTAGAGA AAAGCCCCTG AAAATTGAG AAAACAAACA AGAAACTACT TACCAAGCTAT	10564
TTGAATTGCT GGAATCACAG GCCATTGCTG AGCTGCCTGA ACTGGAAACA CAACAGAAGG	10624
AAAACAAACC ACTCTGATAA TCATTGAGTC AAGTACAGCA GGTGATTGAG GACTGCTGAG	10684
AGGTACAGGC CAAAATTCTT ATGTTGTATT ATAATAATGT CATCTTATAA TACTGTCAGT	10744
ATTTTATAAA ACATTCTCA CAAACTCACA CACATTAAA AACAAAACAC TGTCTCTAAA	10804
ATCCCCAAAT TTTTCATAAA C	10825

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 348 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Gly	Pro	Arg	Ala	Arg	Pro	Ala	Leu	Leu	Leu	Leu	Met	Leu	Leu	Gln
1				5				10				15			
Thr	Ala	Val	Leu	Gln	Gly	Arg	Leu	Leu	Arg	Ser	His	Ser	Leu	His	Tyr
				20				25				30			
Leu	Phe	Met	Gly	Ala	Ser	Glu	Gln	Asp	Leu	Gly	Leu	Ser	Leu	Phe	Glu
				35				40				45			
Ala	Leu	Gly	Tyr	Val	Asp	Asp	Gln	Leu	Phe	Val	Phe	Tyr	Asp	His	Glu

50	55	60
Ser Arg Arg Val Glu Pro Arg Thr Pro Trp Val Ser Ser Arg Ile Ser		
65	70	75
Ser Gln Met Trp Leu Gln Leu Ser Gln Ser Leu Lys Gly Trp Asp His		
85	90	95
Met Phe Thr Val Asp Phe Trp Thr Ile Met Glu Asn His Asn His Ser		
100	105	110
Lys Glu Ser His Thr Leu Gln Val Ile Leu Gly Cys Glu Met Gln Glu		
115	120	125
Asp Asn Ser Thr Glu Gly Tyr Trp Lys Tyr Gly Tyr Asp Gly Gln Asp		
130	135	140
His Leu Glu Phe Cys Pro Asp Thr Leu Asp Trp Arg Ala Ala Glu Pro		
145	150	155
Arg Ala Trp Pro Thr Lys Leu Glu Trp Glu Arg His Lys Ile Arg Ala		
165	170	175
Arg Gln Asn Arg Ala Tyr Leu Glu Arg Asp Cys Pro Ala Gln Leu Gln		
180	185	190
Gln Leu Leu Glu Leu Gly Arg Gly Val Leu Asp Gln Gln Val Pro Pro		
195	200	205
Leu Val Lys Val Thr His His Val Thr Ser Ser Val Thr Thr Leu Arg		
210	215	220
Cys Arg Ala Leu Asn Tyr Tyr Pro Gln Asn Ile Thr Met Lys Trp Leu		
225	230	235
Lys Asp Lys Gln Pro Met Asp Ala Lys Glu Phe Glu Pro Lys Asp Val		
245	250	255
Leu Pro Asn Gly Asp Gly Thr Tyr Gln Gly Trp Ile Thr Leu Ala Val		
260	265	270
Pro Pro Gly Glu Glu Gln Arg Tyr Thr Cys Gln Val Glu His Pro Gly		
275	280	285
Leu Asp Gln Pro Leu Ile Val Ile Trp Glu Pro Ser Pro Ser Gly Thr		
290	295	300
Leu Val Ile Gly Val Ile Ser Gly Ile Ala Val Phe Val Val Ile Leu		
305	310	315
Phe Ile Gly Ile Leu Phe Ile Ile Leu Arg Lys Arg Gln Gly Ser Arg		
325	330	335
Gly Ala Met Gly His Tyr Val Leu Ala Glu Arg Glu		
340	345	

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10825 base pairs
 - (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: join(361..436, 3762..4025, 4235..4510, 5606..5881, 6040..6153, 7107..7147)
- (D) OTHER INFORMATION: /product= "Hereditary Hemochromatosis (HH) protein containing the 24d1 mutation"
/note= "Hereditary Hemochromatosis (HH) gene 24d1 allele"

(ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 140..7319
- (D) OTHER INFORMATION: /note= "start and stop positions for 24d1 allele cDNA (SEQ ID NO:10)"

(ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 3852..3891
- (D) OTHER INFORMATION: /note= "start and stop positions for genomic sequence surrounding variant for 24d2(C) allele (SEQ ID NO:41)"

(ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 5507..6023
- (D) OTHER INFORMATION: /note= "start and stop positions for genomic sequence surrounding variant for 24d1(A) allele (SEQ ID NO:21)"

(ix) FEATURE:

- (A) NAME/KEY: allele
- (B) LOCATION: replace(5834, "a")
- (D) OTHER INFORMATION: /phenotype= "Hereditary Hemochromatosis (HH)"
/label= 24d1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TCTAAGGTTG AGATAAAATT TTTAAATGTA TGATTGAATT TTGAAAATCA TAAATATTTA	60
AATATCTAAA GTTCAGATCA GAACATTGCG AAGCTACTTT CCCCAATCAA CAACACCCCT	120
TCAGGATTAA AAAACCAAGG GGGACACTGG ATCACCTAGT GTTTCACAAG CAGGTACCTT	180
CTGCTGTAGG AGAGAGAGAA CTAAAGTTCT GAAAGACCTG TTGCTTTCA CCAGGAAGTT	240
TTACTGGGCA TCTCCTGAGC CTAGGCAATA GCTGTAGGGT GACTTCTGGA GCCATCCCCG	300
TTTCCCCGCC CCCCAAAAGA AGCGGAGATT TAACGGGGAC GTGCGGCCAG AGCTGGGGAA	360
ATG GGC CCG CGA GCC AGG CCG GCG CTT CTC CTC CTG ATG CTT TTG CAG Met Gly Pro Arg Ala Arg Pro Ala Leu Leu Leu Leu Met Leu Leu Gln	408
1 5 10 15	

ACC GCG GTC CTG CAG GGG CGC TTG CTG C	GTGAGTCCGA	GGGCTGCGGG	456
Thr Ala Val Leu Gln Gly Arg Leu Leu			
20	25		
CGAACTAGGG GCGCGCGGG GGTGGAAAAA TCGAAACTAG	CTTTTCTTT	GCGCTTGGGA	516
GTGGCTAAC TTTGGAGGAC CTGCTCAACC CTATCCGCAA	GCCCCCTCTCC	CTACTTTCTG	576
CGTCCAGACC CCGTGAGGGA GTGCCTACCA CTGAACGTCA	GATAGGGGTC	CCTCGCCCCA	636
GGACCTGCC CCTCCCCCGG CTGTCCCGGC TCTGCGGAGT	GACTTTGGA	ACCGCCCACT	696
CCCTTCCCCC AACTAGAATG CTTTTAAATA AATCTCGTAG	TTCCCTCACTT	GAGCTGAGCT	756
AAGCCTGGGG CTCCTTGAAC CTGGAACTCG GGTTTATTTC	CAATGTCAGC	TGTGCAGTTT	816
TTTCCCCAGT CATCTCCAAA CAGGAAGTTC TTCCCTGAGT	GCTTGCCGAG	AAGGCTGAGC	876
AAACCCACAG CAGGATCCGC ACGGGGTTTC CACCTCAGAA	CGAATGCGTT	GGGCGGTGGG	936
GGCGCAGAAC AGTGGCGTTG GGGATCTGAA TTCTTCACCA	TTCCACCCAC	TTTTGGTGAG	996
ACCTGGGGTG GAGGTCTCTA GGGTGGGAGG CTCCTGAGAG	AGGCCTACCT	CGGGCCTTT	1056
CCCACCTTG GCAATTGTTTC TTTTGCTGG AAAATTAAGT	ATATGTTAGT	TTTGAACGTT	1116
TGAACGTGAAAC AATTCTCTTT TCGGCTAGGC TTTATTGATT	TGCAATGTGC	TGTGTAATTA	1176
AGAGGCCTCT CTACAAAGTA CTGATAATGA ACATGTAAGC	AATGCACTCA	CTTCTAAGTT	1236
ACATTCAATAT CTGATCTTAT TTGATTTCA CTAGGCATAG	GGAGGTAGGA	GCTAATAATA	1296
CGTTTATTTT ACTAGAAGTT AACTGGAATT CAGATTATAT	AACTCTTTTC	AGGTTACAAA	1356
GAACATAAAAT AATCTGGTTT TCTGATGTTA TTTCAAGTAC	TACAGCTGCT	TCTAATCTTA	1416
GTTGACAGTG ATTTTGCCT GTAGTGTAGC ACAGTGTCT	GTGGGTACA	CGCCGGCCTC	1476
AGCACAGCAC TTTGAGTTT GGTACTACGT GTATCCACAT	TTTACACATG	ACAAGAATGA	1536
GGCATGGCAC GGCGCTGCTTC CTGGCAAATT TATTCAATGG	TACACTGGC	TTTGGTGGCA	1596
GAGCTCATGT CTCCACTTCA TAGCTATGAT TCTTAAACAT	CACACTGCAT	TAGAGGTTGA	1656
ATAATAAAAT TTCATGTTGA GCAGAAATAT TCATTGTTA	CAAGTGTAAA	TGAGTCCCAG	1716
CCATGTGTTG CACTGTTCAA GCCCCAAGGG AGAGAGCAGG	GAAACAAGTC	TTTACCCCTT	1776
GATATTTGC ATTCTAGTGG GAGAGATGAC AATAAGCAAA	TGAGCAGAAA	GATATACAAC	1836
ATCAGGAAAT CATGGGTGTT GTGAGAAGCA GAGAAGTCAG	GGCAAGTCAC	TCTGGGCTG	1896
ACACTTGAGC AGAGACATGA AGGAAATAAG AATGATATTG	ACTGGGAGCA	GTATTTCCCA	1956
GGCAAACGTGA GTGGGCCTGG CAAGTTGGAT TAAAAAGCGG	GTTTCTCAG	CACTACTCAT	2016
GTGTGTGTGT GTGGGGGGGG GGGGCGCGT GGGGGTGGGA	AGGGGGACTA	CCATCTGCAT	2076
GTAGGATGTC TAGCACTATC CTGTCCTCCC TACTCACTAG	GTGCTAGGAG	CACTCCCCCA	2136
GTCTTGACAA CCAAAATGT CTCTAAACTT TGCCACATGT	CACCTAGTAG	ACAAACTCCT	2196

GGTTAAGAAG	CTCGGGTTGA	AAAAAATAAA	CAAGTAGTGC	TGGGGAGTAG	AGGCCAAGAA	2256										
GTAGGTAATG	GGCTCAGAAG	AGGAGCCACA	AACAAGGTTG	TGCAGGCGCC	TGTAGGCTGT	2316										
GGTGTGAATT	CTAGCCAAGG	AGTAACAGTG	ATCTGTACCA	GGCTTTAAA	AGATTGCTCT	2376										
GGCTGCTATG	TGGAAAGCAG	AATGAAGGGA	GCAACAGTAA	AAGCAGGGAG	CCCAGCCAGG	2436										
AAGCTGTTAC	ACAGTCCAGG	CAAGAGGTAG	TGGAGTGGGC	TGGGTGGAA	CAGAAAAGGG	2496										
AGTGACAAAC	CATTGTCTCC	TGAATATATT	CTGAAGGAAG	TTGCTGAAGG	ATTCTATGTT	2556										
GTGTGAGAGA	AAGAGAAGAA	TTGGCTGGGT	GTAGTAGCTC	ATGCCAAGGA	GGAGGCCAAG	2616										
GAGAGCAGAT	TCCTGAGCTC	AGGAGTTCAA	GACCAGCCTG	GGCAACACAG	CAAAACCCCT	2676										
TCTCTACAAA	AAATACAAAA	ATTAGCTGGG	TGTGGTGGCA	TGCACCTGTG	ATCCTAGCTA	2736										
CTCGGGAGGC	TGAGGTGGAG	GGTATTGCTT	GAGCCCAGGA	AGTTGAGGCT	GCAGTGAGCC	2796										
ATGACTGTGC	CACTGTACTT	CAGCCTAGGT	GACAGAGCAA	GACCTGTCT	CCCCTGACCC	2856										
CCTGAAAAAG	AGAAGAGTTA	AAGTTGACTT	TGTTCTTTAT	TTTAATTTA	TTGGCCTGAG	2916										
CAGTGGGTA	ATTGGCAATG	CCATTTCTGA	GATGGTGAAG	GCAGAGGAAA	GAGCAGTTTG	2976										
GGGTAAATCA	AGGATCTGCA	TTTGGGACAT	GTAAAGTTG	AGATTCCAGT	CAGGCTTCCA	3036										
AGTGGTGAGG	CCACATAGGC	AGTCAGTGT	AAGAATTCA	GACCAAGGCT	GGGCACGGTG	3096										
GCTCACTTCT	GTAATCCCAG	CACTTTGGTG	GCTGAGGCAG	GTAGATCATT	TGAGGTCAAG	3156										
AGTTTGAGAC	AAGCTTGGCC	AACATGGTGA	AACCCATGT	CTACTAAAAA	TACAAAAAATT	3216										
AGCCTGGTGT	GGTGGCGCAC	GCCTATAGTC	CCAGGTTTTC	AGGAGGCTTA	GGTAGGGAGAA	3276										
TCCCTTGAAC	CCAGGAGGTG	CAGGTTGCAG	TGAGCTGAGA	TTGTGCCACT	GCACTCCAGC	3336										
CTGGGTGATA	GAGTGAGACT	CTGTCTAAA	AAAAAAAAAA	AAAAAAAAAA	AAAAAAACTGA	3396										
AGGAATTATT	CCTCAGGATT	TGGGTCTAAT	TTGCCCTGAG	CACCAACTCC	TGAGTTCAAC	3456										
TACCATGGCT	AGACACACCT	TAACATTTC	TAGAATCCAC	CAGCTTAGT	GGAGTCTGTGTC	3516										
TAATCATGAG	TATTGGAATA	GGATCTGGGG	GCAGTGAGGG	GGTGGCAGCC	ACGTGTGGCA	3576										
GAGAAAAGCA	CACAAGGAAA	GAGCACCCAG	GACTGTCATA	TGGAAGAAAG	ACAGGACTGC	3636										
AACTCACCCCT	TCACAAAATG	AGGACCAGAC	ACAGCTGATG	GTATGAGTTG	ATGCAGGTGT	3696										
GTGGAGCCTC	AACATCCTGC	TCCCCCTCTA	CTACACATGG	TTAAGGCCTG	TTGCTCTGTC	3756										
TCCAG	GT	TCA	CAC	TCT	CTG	CAC	TAC	CTC	TTC	ATG	GGT	GCC	TCA	GAG	3802	
Arg	Ser	His	Ser	Leu	His	Tyr	Leu	Phe	Met	Gly	Ala	Ser	Glu			
30														35		
CAG	GAC	CTT	GGT	CTT	TCC	TTG	TTT	GAA	GCT	TTG	GGC	TAC	GTG	GAT	GAC	3850
Gln	Asp	Leu	Gly	Leu	Ser	Leu	Phe	Glu	Ala	Leu	Gly	Tyr	Val	Asp	Asp	
40															55	
CAG	CTG	TTC	GTG	TTC	TAT	GAT	CAT	GAG	AGT	CGC	CGT	GTG	GAG	CCC	CGA	3898
Gln	Leu	Phe	Val	Phe	Tyr	Asp	His	Glu	Ser	Arg	Arg	Val	Glu	Pro	Arg	

	60	65	70	
ACT CCA TGG GTT TCC AGT AGA ATT TCA AGC CAG ATG TGG CTG CAG CTG Thr Pro Trp Val Ser Ser Arg Ile Ser Ser Gln Met Trp Leu Gln Leu	75	80	85	3946
AGT CAG AGT CTG AAA GGG TGG GAT CAC ATG TTC ACT GTT GAC TTC TGG Ser Gln Ser Leu Lys Gly Trp Asp His Met Phe Thr Val Asp Phe Trp	90	95	100	3994
ACT ATT ATG GAA AAT CAC AAC CAC AGC AAG G GTATGTGGAG AGGGGGCCTC Thr Ile Met Glu Asn His Asn His Ser Lys	105	110		4045
ACCTTCCTGA GGTTGTCAGA GCTTTTCATC TTTTCATGCA TCTTGAAGGA AACAGCTGGA				4105
AGTCTGAGGT CTTGTGGGAG CAGGGAAAGAG GGAAGGAATT TGCTTCCTGA GATCATTGG				4165
TCCTTGGGGA TGGTGGAAAT AGGGACCTAT TCCTTGGTT GCAGTTAACCA AGGCTGGGGA				4225
TTTTTCCAG AG TCC CAC ACC CTG CAG GTC ATC CTG GGC TGT GAA ATG Glu Ser His Thr Leu Gln Val Ile Leu Gly Cys Glu Met	115	120	125	4272
CAA GAA GAC AAC AGT ACC GAG GGC TAC TGG AAG TAC GGG TAT GAT GGG Gln Glu Asp Asn Ser Thr Glu Gly Tyr Trp Lys Tyr Gly Tyr Asp Gly	130	135	140	4320
CAG GAC CAC CTT GAA TTC TGC CCT GAC ACA CTG GAT TGG AGA GCA GCA Gln Asp His Leu Glu Phe Cys Pro Asp Thr Leu Asp Trp Arg Ala Ala	145	150	155	4368
GAA CCC AGG GCC TGG CCC ACC AAG CTG GAG TGG GAA AGG CAC AAG ATT Glu Pro Arg Ala Trp Pro Thr Lys Leu Glu Trp Glu Arg His Lys Ile	160	165	170	4416
CGG GCC AGG CAG AAC AGG GCC TAC CTG GAG AGG GAC TGC CCT GCA CAG Arg Ala Arg Gln Asn Arg Ala Tyr Leu Glu Arg Asp Cys Pro Ala Gln	175	180	185	4464
CTG CAG CAG TTG CTG GAG CTG GGG AGA GGT GTT TTG GAC CAA CAA G Leu Gln Gln Leu Glu Leu Gly Arg Gly Val Leu Asp Gln Gln	195	200	205	4510
GTATGGTGGAA AACACACTTC TGCCCCTATA CTCTAGTGGC AGAGTGGAGG AGGTTGCAGG				4570
GCACGGAATC CCTGGTTGGAA GTTTCAGAGG TGGCTGAGGC TGTGTGCCTC TCCAAATTCT				4630
GGGAAGGGAC TTTCTCAATC CTAGAGTCTC TACCTTATAA TTGAGATGTA TGAGACAGCC				4690
ACAAGTCATG GGTTTAATTT CTTTTCTCCA TGCATATGGC TCAAAGGGAA GTGTCTATGG				4750
CCCTTGCTTT TTATTTAACCC AATAATCTTT TGTATATTAA TACCTGTTAA AAATTCAGAA				4810
ATGTCAAGGC CGGGCACGGT GGCTCACCCCC TGTAATCCCA GCACTTTGGG AGGCCGAGGC				4870
GGGTGGTCAC AAGGTCAGGA GTTTGAGACC AGCCTGACCA ACATGGTAA ACCCGTCTCT				4930
AAAAAAATAC AAAAATTAGC TGGTCACAGT CATGCGCACC TGTAGTCCCA GCTAATTGGA				4990
AGGCTGAGGC AGGAGCATCG CTTGAACCTG GGAAGCGGAA GTTGCAGTGA GCCAAGATCG				5050

CGCCACTGCA CTCCAGCCTA GGCAGCAGAG TGAGACTCCA TCTTAAAAAA AAAAAAAA	5110
AAAAAAAGAG AATTCAAGAGA TCTCAGCTAT CATATGAATA CCAGGACAAA ATATCAAGTG	5170
AGGCCACTTA TCAGAGTAGA AGAACCTTT AGGTTAAAG TTTCTTCAT AGAACATAGC	5230
AATAATCACT GAAGCTACCT ATCTTACAAG TCCGCTTCTT ATAACAATGC CTCCTAGGTT	5290
GACCCAGGTG AAACTGACCA TCTGTATTCA ATCATTTCA ATGCACATAA AGGGCAATT	5350
TATCTATCAG AACAAAGAAC ATGGGTAACA GATATGTATA TTTACATGTG AGGAGAACAA	5410
GCTGATCTGA CTGCTCTCCA AGTGACACTG TGTTAGAGTC CAATCTTAGG ACACAAAATG	5470
GTGTCTCTCC TGTAGCTTGT TTTTTCTGA AAAGGGTATT TCCTTCCTCC AACCTATAGA	5530
AGGAAGTGAA AGTTCCAGTC TTCCTGGCAA GGGTAAACAG ATCCCCTCTC CTCATCCTTC	5590
CTCTTCCTG TCAAG TG CCT CCT TTG GTG AAG GTG ACA CAT CAT GTG ACC	5640
Val Pro Pro Leu Val Lys Val Thr His His Val Thr	
210 215	
TCT TCA GTG ACC ACT CTA CGG TGT CGG GCC TTG AAC TAC TAC CCC CAG	5688
Ser Ser Val Thr Thr Leu Arg Cys Arg Ala Leu Asn Tyr Tyr Pro Gln	
220 225 230	
AAC ATC ACC ATG AAG TGG CTG AAG GAT AAG CAG CCA ATG GAT GCC AAG	5736
Asn Ile Thr Met Lys Trp Leu Lys Asp Lys Gln Pro Met Asp Ala Lys	
235 240 245	
GAG TTC GAA CCT AAA GAC GTA TTG CCC AAT GGG GAT GGG ACC TAC CAG	5784
Glu Phe Glu Pro Lys Asp Val Leu Pro Asn Gly Asp Gly Thr Tyr Gln	
250 255 260 265	
GGC TGG ATA ACC TTG GCT GTA CCC CCT GGG GAA GAG CAG AGA TAT ACG	5832
Gly Trp Ile Thr Leu Ala Val Pro Pro Gly Glu Glu Gln Arg Tyr Thr	
270 275 280	
TAC CAG GTG GAG CAC CCA GGC CTG GAT CAG CCC CTC ATT GTG ATC TGG G	5881
Tyr Gln Val Glu His Pro Gly Leu Asp Gln Pro Leu Ile Val Ile Trp	
285 290 295	
GTATGTGACT GATGAGAGCC AGGAGCTGAG AAAATCTATT GGGGGTTGAG AGGAGTGCCT	5941
GAGGAGGTAA TTATGGCAGT GAGATGAGGA TCTGCTCTT GTTAGGGGT GGGCTGAGGG	6001
TGGCAATCAA AGGCTTTAAC TTGCTTTTC TGTTTAG AG CCC TCA CCG TCT	6053
Glu Pro Ser Pro Ser	
300	
GGC ACC CTA GTC ATT GGA GTC ATC AGT GGA ATT GCT GTT TTT GTC GTC	6101
Gly Thr Leu Val Ile Gly Val Ile Ser Gly Ile Ala Val Phe Val Val	
305 310 315	
ATC TTG TTC ATT GGA ATT TTG TTC ATA ATA TTA AGG AAG AAG AGG CAG GGT	6149
Ile Leu Phe Ile Gly Ile Leu Phe Ile Ile Leu Arg Lys Arg Gln Gly	
320 325 330	
TCA A GTGAGTAGGA ACAAGGGGAA AGTCTCTTAG TACCTCTGCC CCAGGGCACA	6203
Ser	
335	

GTGGGAAGAG GGGCAGAGGG GATCTGGCAT CCATGGGAAG CATTTCCTC ATTTATATTC	6263
TTTGGGGACA CCAGCAGCTC CCTGGGAGAC AGAAAATAAT GGTTCTCCC AGAATGAAAG	6323
TCTCTAATTC AACAAACATC TTCAGAGCAC CTACTATTT GCAAGAGCTG TTTAAGGTAG	6383
TACAGGGCT TTGAGGTTGA GAAGTCACTG TGGCTATTCT CAGAACCAA ATCTGGTAGG	6443
GAATGAAATT GATAGCAAGT AAATGTAGTT AAAGAAGACC CCATGAGGTC CTAAAGCAGG	6503
CAGGAAGCAA ATGCCTAGGG TGTCAAAGGA AAGAATGATC ACATTCAGCT GGGGATCAAG	6563
ATAGCCTTCT GGATCTTGAA GGAGAAGCTG GATTCCATTA GGTGAGGTTG AAGATGATGG	6623
GAGGTCTACA CAGACGGAGC AACCATGCCA AGTAGGAGAG TATAAGGCAT ACTGGGAGAT	6683
TAGAAATAAT TACTGTACCT TAACCCTGAG TTTGCGTAGC TATCACTCAC CAATTATGCA	6743
TTTCTACCCC CTGAACATCT GTGGTGTAGG GAAAAGAGAA TCAGAAAGAA GCCAGCTCAT	6803
ACAGAGTCCA AGGGCTTTT GGGATATTGG GTTATGATCA CTGGGGTGT ATTGAAGGAT	6863
CCTAAGAAAG GAGGACCACG ATCTCCCTTA TATGGTGAAT GTGTTGTTAA GAAGTTAGAT	6923
GAGAGGTGAG GAGACCAGTT AGAAAGCCAA TAAGCATTTC CAGATGAGAG ATAATGGTTC	6983
TTGAAATCCA ATAGTGCCC A GTCTAAATT GAGATGGGTG AATGAGGAAA ATAAGGAAGA	7043
GAGAAGAGGC AAGATGGTGC CTAGGTTGT GATGCCTCTT TCCTGGGTCT CTTGTCTCCA	7103
CAG GA GGA GCC ATG GGG CAC TAC GTC TTA GCT GAA CGT GAG	7144
Arg Gly Ala Met Gly His Tyr Val Leu Ala Glu Arg Glu	
340 345	
TGACACGCAG CCTGCAGACT CACTGTGGGA AGGAGACAAA ACTAGAGACT CAAAGAGGGA	7204
GTGCATTTAT GAGCTCTTCA TGTTCAGGA GAGAGTTGAA CCTAAACATA GAAATTGCCT	7264
GACGAACCTC TTGATTTAG CCTTCTCTGT TCATTTCTC AAAAAGATTT CCCCATTTAG	7324
GTTTCTGAGT TCCTGCATGC CGGTGATCCC TAGCTGTGAC CTCTCCCTG GAACTGTCTC	7384
TCATGAACCT CAAGCTGCAT CTAGAGGCTT CCTTCATTTC CTCCGTCACC TCAGAGACAT	7444
ACACCTATGT CATTCTATT CCTATTTTG GAAGAGGACT CCTTAAATTT GGGGGACTTA	7504
CATGATTCAT TTTAACATCT GAGAAAAGCT TTGAACCTG GGACGTGGCT AGTCATAACC	7564
TTACCAGATT TTTACACATG TATCTATGCA TTTCTGGAC CCGTTCAACT TTTCTTTGA	7624
ATCCTCTCTC TGTGTTACCC AGTAACTCAT CTGTCACCAA GCCTGGGGA TTCTCCATC	7684
TGATTGTGAT GTGAGTTGCA CAGCTATGAA GGCTGTACAC TGCACGAATG GAAGAGGCAC	7744
CTGTCCCAGA AAAAGCATCA TGGCTATCTG TGGTAGTAT GATGGGTGTT TTTAGCAGGT	7804
AGGAGGCAA TATCTTGAAA GGGGTTGTGA AGAGGTGTTT TTTCTAATTG GCATGAAGGT	7864
GTCATACAGA TTTGCAAAGT TTAATGGTGC CTTCATTGG GATGCTACTC TAGTATTCCA	7924
GACCTGAAGA ATCACAATAA TTTCTACCT GGTCTCTCCT TGTGTTGATA ATGAAAATTA	7984

TGATAAGGAT GATAAAAGCA CTTACTTCGT GTCCGACTCT TCTGAGCACC TACTTACATG	8044
CATTACTGCA TGCACTTCTT ACAATAATTC TATGAGATAG GTACTATTAT CCCCATTCT	8104
TTTTAAATG AAGAAAGTGA AGTAGGCCGG GCACGGTGGC TCACGCCTGT AATCCCAGCA	8164
CTTTGGGAGG CCAAAGCGGG TGGATCACGA GGTCAAGGAGA TCGAGACCAT CCTGGCTAAC	8224
ATGGTAAAC CCCATCTCTA ATAAAAATAC AAAAAAATTAG CTGGCGTGG TGGCAGACGC	8284
CTGTAGTCCC AGCTACTCGG AAGGCTGAGG CAGGAGAATG GCATGAACCC AGGAGGCAGA	8344
GCTTGCAGTG AGCCGAGTTT GCGCCACTGC ACTCCAGCCT AGGTGACAGA GTGAGACTCC	8404
ATCTCAAAA AATAAAAATA AAAATAAAA AATGAAAAAA AAAAGAAAGT GAAGTATAGA	8464
GTATCTCATA GTTGTCAAGT GATAGAAACA GGTTCAAAC TCAGTCAATC TGACCGTTG	8524
ATACATCTCA GACACCACTA CATTCACTAG TTTAGATGCC TAGAATAAT AGAGAAGGAA	8584
GGAGATGGCT CTTCTCTGT CTCATTGTGT TTCTCTGAG TGAGCTTGAA TCACATGAAG	8644
GGGAACAGCA GAAAACAACC AACTGATCCT CAGCTGTCAT GTTCCCTTA AAAGTCCCTG	8704
AAGGAAGGTC CTGGAATGTG ACTCCCTTGC TCCTCTGTTG CTCTCTTGG CATTCAATTG	8764
TTTGGACCT ACACGGTGT TGTAATTGGT GGGGACAGCT AGTGGCCCTG CTGGGCTTCA	8824
CACACGGTGT CCTCCCTAGG CCAGTGCCTC TGGAGTCAGA ACTCTGGTGG TATTTCCCTC	8884
AATGAAGTGG AGTAAGCTCT CTCATTTGA GATGGTATAA TGGAAGCCAC CAAGTGGCTT	8944
AGAGGATGCC CAGGTCCCTC CATGGAGCCA CTGGGGTTCC GGTGCACATT AAAAAAAA	9004
TCTAACCAAGG ACATTCAGGA ATTGCTAGAT TCTGGAAAT CAGTTCACCA TGTTCAAAAG	9064
AGTCTTTTT TTTTTTTGTA GACTCTATTG CCCAGGCTGG AGTCAATGG CATGATCTCG	9124
GCTCACTGTA ACCTCTGCCT CCCAGGTTCA AGCGATTCTC CTGTCTCAGC CTCCCAAGTA	9184
GCTGGGATTA CAGGCCTGCA CCACCATGCC CGGCTAATTT TTGTATTTT AGTAGAGACA	9244
GGGTTTCACC ATGTTGGCCA GGCTGGCTC GAACTCTCCT GACCTCGTGA TCCGCCTGCC	9304
TCGGCCTCCC AAAGTGCTGA GATTACAGGT GTGAGCCACC CTGCCAGCC GTCAAAAGAG	9364
TCTTAATATA TATATCCAGA TGGCATGTGT TTACTTTATG TTACTACATG CACTTGGCTG	9424
CATAAAATGTG GTACAAGCAT TCTGTCTTGA AGGCAGGTG CTTCAAGGATA CCATATACAG	9484
CTCAGAAGTT TCTCTTTAG GCATTAATT TTAGCAAAGA TATCTCATCT CTTCTTTAA	9544
ACCATTTCT TTTTTGTGG TTAGAAAAGT TATGTAGAAA AAAGTAAATG TGATTTACGC	9604
TCATTGTAGA AAAGCTATAA AATGAATACA ATTAAAGCTG TTATTTAATT AGCCAGTGAA	9664
AAACTATTAA CAACTGTCT ATTACCTGTT AGTATTATTG TTGCATTAAG AATGCATATA	9724
CTTTAATAAA TGTATATTGT ATTGTATACT GCATGATTTC ATTGAAGTTC TTGTTCATCT	9784
TGTGTATATA CTTAATCGCT TTGTCATTT GGAGACATTG ATTTGCTTC TAATTTCTT	9844

ACATTTGTC TTACGGAATA TTTTCATTCA ACTGTGGTAG CCGAATTAAT CGTGTTCCTT	9904
CACTCTAGGG ACATTGTCGT CTAAGTTGTA AGACATTGGT TATTTACCA GCAAACCATT	9964
CTGAAAGCAT ATGACAAATT ATTTCTCTCT TAATATCTTA CTATACTGAA AGCAGACTGC	10024
TATAAGGCTT CACTTACTCT TCTACCTCAT AAGGAATATG TTACAATTAA TTTATTAGGT	10084
AAGCATTGT TTTATATTGG TTTTATTCA CCTGGGCTGA GATTCAAGA AACACCCCAG	10144
TCTTCACAGT AACACATTTC ACTAACACAT TTACTAAACA TCAGCAACTG TGGCCTGTTA	10204
ATTTTTTAA TAGAAATTAA AAGTCCTCAT TTTCTTCGG TGTTTTTAA GCTTAATTAA	10264
TCTGGCTTTA TTCATAAAATT CTTAAGGTCA ACTACATTG AAAAATCAAA GACCTGCATT	10324
TTAAATTCTT ATTCACCTCT GGCAAAACCA TTCACAAACC ATGGTAGTAA AGAGAAGGGT	10384
GACACCTGGT GGCCATAGGT AAATGTACCA CGGTGGTCCG GTGACCAGAG ATGCAGCGCT	10444
GAGGGTTTTC CTGAAGGTAA AGGAATAAAG AATGGGTGGA GGGGCGTGCA CTGGAAATCA	10504
CTTGTAGAGA AAAGCCCCTG AAAATTGAG AAAACAAACA AGAAACTACT TACCAGCTAT	10564
TTGAATTGCT GGAATCACAG GCCATTGCTG AGCTGCCTGA ACTGGGAACA CAACAGAAGG	10624
AAAACAAACC ACTCTGATAA TCATTGAGTC AAGTACAGCA GGTGATTGAG GACTGCTGAG	10684
AGGTACAGGC CAAAATTCTT ATGTTGTATT ATAATAATGT CATCTTATAA TACTGTCAGT	10744
ATTTTATAAA ACATTCTTCA CAAACTCACA CACATTAAA AACAAAACAC TGTCTCTAAA	10804
ATCCCCAAAT TTTTCATAAA C	10825

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 348 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met	Gly	Pro	Arg	Ala	Arg	Pro	Ala	Leu	Leu	Leu	Leu	Leu	Leu	Gln
1														15

Thr	Ala	Val	Leu	Gln	Gly	Arg	Leu	Leu	Arg	Ser	His	Ser	Leu	His	Tyr
															30
20															

Leu	Phe	Met	Gly	Ala	Ser	Glu	Gln	Asp	Leu	Gly	Leu	Ser	Leu	Phe	Glu
															45
35															

Ala	Leu	Gly	Tyr	Val	Asp	Asp	Gln	Leu	Phe	Val	Phe	Tyr	Asp	His	Glu
															50
50															

Ser	Arg	Arg	Val	Glu	Pro	Arg	Thr	Pro	Trp	Val	Ser	Ser	Arg	Ile	Ser
															60
65															

Ser	Gln	Met	Trp	Leu	Gln	Leu	Ser	Gln	Ser	Leu	Lys	Gly	Trp	Asp	His
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

85

90

95

Met Phe Thr Val Asp Phe Trp Thr Ile Met Glu Asn His Asn His Ser
 100 105 110

Lys Glu Ser His Thr Leu Gln Val Ile Leu Gly Cys Glu Met Gln Glu
 115 120 125

Asp Asn Ser Thr Glu Gly Tyr Trp Lys Tyr Gly Tyr Asp Gly Gln Asp
 130 135 140

His Leu Glu Phe Cys Pro Asp Thr Leu Asp Trp Arg Ala Ala Glu Pro
 145 150 155 160

Arg Ala Trp Pro Thr Lys Leu Glu Trp Glu Arg His Lys Ile Arg Ala
 165 170 175

Arg Gln Asn Arg Ala Tyr Leu Glu Arg Asp Cys Pro Ala Gln Leu Gln
 180 185 190

Gln Leu Leu Glu Leu Gly Arg Gly Val Leu Asp Gln Gln Val Pro Pro
 195 200 205

Leu Val Lys Val Thr His His Val Thr Ser Ser Val Thr Thr Leu Arg
 210 215 220

Cys Arg Ala Leu Asn Tyr Tyr Pro Gln Asn Ile Thr Met Lys Trp Leu
 225 230 235 240

Lys Asp Lys Gln Pro Met Asp Ala Lys Glu Phe Glu Pro Lys Asp Val
 245 250 255

Leu Pro Asn Gly Asp Gly Thr Tyr Gln Gly Trp Ile Thr Leu Ala Val
 260 265 270

Pro Pro Gly Glu Glu Gln Arg Tyr Thr Tyr Gln Val Glu His Pro Gly
 275 280 285

Leu Asp Gln Pro Leu Ile Val Ile Trp Glu Pro Ser Pro Ser Gly Thr
 290 295 300

Leu Val Ile Gly Val Ile Ser Gly Ile Ala Val Phe Val Val Ile Leu
 305 310 315 320

Phe Ile Gly Ile Leu Phe Ile Ile Leu Arg Lys Arg Gln Gly Ser Arg
 325 330 335

Gly Ala Met Gly His Tyr Val Leu Ala Glu Arg Glu
 340 345

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10825 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS
 (B) LOCATION: join(361..436, 3762..4025, 4235..4510, 5606..5881,
 6040..6153, 7107..7147)
 (D) OTHER INFORMATION: /product= "Hereditary Hemochromatosis
 (HH) protein containing the 24d2
 mutation"
 /note= "Hereditary Hemochromatosis (HH)
 gene 24d2 allele"

(ix) FEATURE:

(A) NAME/KEY: -
 (B) LOCATION: 140..7319
 (D) OTHER INFORMATION: /note= "start and stop positions for
 24d2 allele cDNA (SEQ ID NO:11)"

(ix) FEATURE:

(A) NAME/KEY: -
 (B) LOCATION: 3852..3891
 (D) OTHER INFORMATION: /note= "start and stop positions for
 genomic sequence surrounding variant
 for 24d2(G) allele (SEQ ID NO:42)"

(ix) FEATURE:

(A) NAME/KEY: -
 (B) LOCATION: 5507..6023
 (D) OTHER INFORMATION: /note= "start and stop positions for
 genomic sequence surrounding variant
 for 24d1(G) allele (SEQ ID NO:20)"

(ix) FEATURE:

(A) NAME/KEY: allele
 (B) LOCATION: replace(3872, "g")
 (D) OTHER INFORMATION: /phenotype= "Hereditary Hemochromatosis
 (HH)"
 /label= 24d2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TCTAAGGTG AGATAAAATT TTTAAATGTA TGATTGAATT TTGAAAATCA TAAATATTAA	60
AATATCTAAA GTTCAGATCA GAACATTGCG AAGCTACTTT CCCCCAATCAA CAACACCCCT	120
TCAGGATTAA AAAACCAAGG GGGACACTGG ATCACCTAGT GTTTCACAAG CAGGTACCTT	180
CTGCTGTAGG AGAGAGAGAA CTAAAGTTCT GAAAGACCTG TTGCTTTCA CCAGGAAGTT	240
TTACTGGGCA TCTCCTGAGC CTAGGCAATA GCTGTAGGGT GACTTCTGGA GCCATCCCCG	300
TTTCCCCGCC CCCCAAAAGA AGCGGAGATT TAACGGGGAC GTGCGGCCAG AGCTGGGGAA	360
ATG GGC CCG CGA GCC AGG CCG GCG CTT CTC CTC CTG ATG CTT TTG CAG Met Gly Pro Arg Ala Arg Pro Ala Leu Leu Leu Leu Met Leu Leu Gln	408
1 5 10 15	
ACC GCG GTC CTG CAG GGG CGC TTG CTG C GTGAGTCCGA GGGCTGCGGG Thr Ala Val Leu Gln Gly Arg Leu Leu	456
20 25	
CGAACTAGGG GCGCGGCGGG GGTGGAAAAA TCGAAACTAG CTTTTCTTT GCGCTTGGGA	516

GTGGCTAAC	TTGGAGGAC	CTGCTAAC	CTATCCGCAA	GCCCCTCTCC	CTACTTCTG	576
CGTCCAGACC	CCGTGAGGGA	GTGCCTACCA	CTGAACGTCA	GATAGGGGTC	CCTCGCCCCA	636
GGACCTGCC	CCTCCCCCGG	CTGTCCCGGC	TCTGCGGAGT	GACTTTGGA	ACCGCCCAC	696
CCCTTCCCCC	AACTAGAATG	CTTTAAATA	AATCTCGTAG	TTCTCACTT	GAGCTGAGCT	756
AAGCCTGGGG	CTCCTTGAAC	CTGGAAC	CTCG	GGTTTATTTC	CAATGTCAGC	TGTGCAGTT
TTTCCCCAGT	CATCTCCAAA	CAGGAAGTTC	TTCCCTGAGT	GCTTGCCGAG	AAGGCTGAGC	816
AAACCCACAG	CAGGATCCGC	ACGGGGTTTC	CACCTCAGAA	CGAATGCGTT	GGGCGGTGGG	936
GGCGCGAAAG	AGTGGCGTTG	GGGATCTGAA	TTCTTCACCA	TTCCACCCAC	TTTTGGTGAG	996
ACCTGGGGTG	GAGGTCTCTA	GGGTGGGAGG	CTCCTGAGAG	AGGCCTACCT	CGGGCCTTTC	1056
CCCACTCTTG	GCAATTGTTTC	TTTGCCTGG	AAAATTAAGT	ATATGTTAGT	TTGAACGTT	1116
TGAAC	AATTCTCTTT	TCGGCTAGGC	TTTATTGATT	TGCAATGTGC	TGTGTAATTA	1176
AGAGGCCTCT	CTACAAAGTA	CTGATAATGA	ACATGTAAGC	AATGCACTCA	CTTCTAAGTT	1236
ACATTCA	ATCTGATCTTAT	TTGATTTCA	CTAGGCATAG	GGAGGTAGGA	GCTAATAATA	1296
CGTTTATTTT	ACTAGAAGTT	AACTGGAATT	CAGATTATAT	AACTCTTTC	AGGTTACAAA	1356
GAACATAAAAT	AATCTGGTTT	TCTGATGTTA	TTTCAAGTAC	TACAGCTGCT	TCTAATCTTA	1416
GTTGACAGTG	ATTTTGCCT	GTAGTGTAGC	ACAGTGTCT	GTGGGTACAC	CGCCGGCCTC	1476
AGCACAGCAC	TTTGAGTTTT	GGTACTACGT	GTATCCACAT	TTTACACATG	ACAAGAATGA	1536
GGCATGGCAC	GGCCTGCTTC	CTGGCAAATT	TATTCAATGG	TACACTGGGC	TTTGGTGGCA	1596
GAGCTCATGT	CTCCACTTCA	TAGCTATGAT	TCTTAAACAT	CACACTGCAT	TAGAGGTTGA	1656
ATAATAAAAT	TTCATGTTGA	GCAGAAATAT	TCATTGTTA	CAAGTGTAAA	TGAGTCCCAG	1716
CCATGTGTTG	CACTGTTCAA	GCCCCAAGGG	AGAGAGCAGG	GAAACAAGTC	TTTACCCCTT	1776
GATATTTGC	ATTCTAGTGG	GAGAGATGAC	AATAAGCAA	TGAGCAGAAA	GATATACAAC	1836
ATCAGGAAAT	CATGGGTGTT	GTGAGAAGCA	GAGAAGTCAG	GGCAAGTCAC	TCTGGGGCTG	1896
ACACTTGAGC	AGAGACATGA	AGGAAATAAG	AATGATATTG	ACTGGGAGCA	GTATTTCCCA	1956
GGCAAAC	GTGGGCCTGG	CAAGTTGGAT	AAAAAGCGG	GTTTTCTCAG	CACTACTCAT	2016
GTGTGTGTGT	GTGGGGGGGG	GGGGCGGC	GGGGGTGGGA	AGGGGGACTA	CCATCTGCAT	2076
GTAGGATGTC	TAGCAGTATC	CTGTCCTCCC	TACTCACTAG	GTGCTAGGAG	CACTCCCCCA	2136
GTCTTGACAA	CCAAAAATGT	CTCTAAACTT	TGCCACATGT	CACCTAGTAG	ACAAACTCCT	2196
GGTTAAGAAG	CTCGGGTTGA	AAAAAATAAA	CAAGTAGTGC	TGGGGAGTAG	AGGCCAAGAA	2256
GTAGGTAATG	GGCTCAGAAG	AGGAGCCACA	AACAAGGTTG	TGCAGGCGCC	TGTAGGCTGT	2316
GGTGTGAATT	CTAGCCAAGG	AGTAACAGTG	ATCTGTCACA	GGCTTTAAA	AGATTGCTCT	2376

GGCTGCTATG TGGAAAGCAG AATGAAGGGA GCAACAGTAA AAGCAGGGAG CCCAGCCAGG	2436		
AAGCTGTTAC ACAGTCCAGG CAAGAGGTAG TGGAGTGGC TGGGTGGAA CAGAAAAGGG	2496		
AGTACAAAC CATTGTCTCC TGAATATATT CTGAAGGAAG TTGCTGAAGG ATTCTATGTT	2556		
GTGTGAGAGA AAGAGAAGAA TTGGCTGGGT GTAGTAGCTC ATGCCAAGGA GGAGGCCAAG	2616		
GAGAGCAGAT TCCTGAGCTC AGGAGTTCAA GACCAGCCTG GGCAACACAG CAAAACCCCT	2676		
TCTCTACAAA AAATACAAAA ATTAGCTGGG TGTGGTGGCA TGCACCTGTG ATCCTAGCTA	2736		
CTCGGGAGGC TGAGGTGGAG GGTATTGCTT GAGCCCAGGA AGTTGAGGCT GCAGTGAGCC	2796		
ATGACTGTGC CACTGTACTT CAGCCTAGGT GACAGAGCAA GACCCTGTCT CCCCTGACCC	2856		
CCTGAAAAAG AGAAGAGTTA AAGTTGACTT TGTTCTTAT TTTAATTAA TTGGCCTGAG	2916		
CAGTGGGGTA ATTGGCAATG CCATTTCTGA GATGGTGAAG GCAGAGGAAA GAGCAGTTG	2976		
GGTAAATCA AGGATCTGCA TTTGGGACAT GTTAAGTTG AGATTCCAGT CAGGCTTCCA	3036		
AGTGGTGAGG CCACATAGGC AGTTCAGTGT AAGAATTCA GACCAAGGCT GGGCACGGTG	3096		
GCTCACTTCT GTAATCCCAG CACTTTGGTG GCTGAGGCAG GTAGATCATT TGAGGTCAGG	3156		
AGTTTGAGAC AAGCTTGGCC AACATGGTGA AACCCATGT CTACTAAAAA TACAAAAAATT	3216		
AGCCTGGTGT GGTGGCGCAC GCCTATAGTC CCAGGTTTC AGGAGGCTTA GGTAGGAGAA	3276		
TCCCTTGAAC CCAGGAGGTG CAGGTTGCAG TGAGCTGAGA TTGTGCCACT GCACTCCAGC	3336		
CTGGGTGATA GAGTGAGACT CTGTCTAAA AAAAAAAA AAAAAAAA AAAAAACTGA	3396		
AGGAATTATT CCTCAGGATT TGGGTCTAAT TTGCCCTGAG CACCAACTCC TGAGTTCAAC	3456		
TACCATGGCT AGACACACCT TAACATTTTC TAGAATCCAC CAGCTTTAGT GGAGTCTGTC	3516		
TAATCATGAG TATTGGAATA GGATCTGGGG GCAGTGAGGG GGTGGCAGCC ACGTGTGGCA	3576		
GAGAAAAGCA CACAAGGAAA GAGCACCCAG GACTGTCATA TGGAAGAAAG ACAGGACTGC	3636		
AACTCACCCCT TCACAAAATG AGGACCAGAC ACAGCTGATG GTATGAGTTG ATGCAGGTGT	3696		
GTGGAGCCTC AACATCCTGC TCCCCTCCTA CTACACATGG TTAAGGCCCTG TTGCTCTGTC	3756		
TCCAG GT TCA CAC TCT CTG CAC TAC CTC TTC ATG GGT GCC TCA GAG	3802		
Arg Ser His Ser Leu His Tyr Leu Phe Met Gly Ala Ser Glu			
30	35		
CAG GAC CTT GGT CTT TCC TTG TTT GAA GCT TTG GGC TAC GTG GAT GAC	3850		
Gln Asp Leu Gly Leu Ser Leu Phe Glu Ala Leu Gly Tyr Val Asp Asp			
40	45	50	55
CAG CTG TTC GTG TTC TAT GAT GAT GAG AGT CGC CGT GTG GAG CCC CGA	3898		
Gln Leu Phe Val Phe Tyr Asp Asp Glu Ser Arg Arg Val Glu Pro Arg			
60	65	70	
ACT CCA TGG GTT TCC AGT AGA ATT TCA AGC CAG ATG TGG CTG CAG CTG	3946		
Thr Pro Trp Val Ser Ser Arg Ile Ser Ser Gln Met Trp Leu Gln Leu			
75	80	85	

AGT CAG AGT CTG AAA GGG TGG GAT CAC ATG TTC ACT GTT GAC TTC TGG Ser Gln Ser Leu Lys Gly Trp Asp His Met Phe Thr Val Asp Phe Trp 90 95 100	3994
ACT ATT ATG GAA AAT CAC AAC CAC AGC AAG G GTATGTGGAG AGGGGGCCTC Thr Ile Met Glu Asn His Asn His Ser Lys 105 110	4045
ACCTTCCTGA GGTTGTCAGA GCTTTTCATC TTTTCATGCA TCTTGAAGGA AACAGCTGGA	4105
AGTCTGAGGT CTTGTGGAG CAGGGAAAGAG GGAAGGAATT TGCTTCCTGA GATCATTGG	4165
TCCTTGGGGA TGGTGGAAAT AGGGACCTAT TCCTTGGTT GCAGTTAACCA AGGCTGGGA	4225
TTTTTCCAG AG TCC CAC ACC CTG CAG GTC ATC CTG GGC TGT GAA ATG Glu Ser His Thr Leu Gln Val Ile Leu Gly Cys Glu Met 115 120 125	4272
CAA GAA GAC AAC AGT ACC GAG GGC TAC TGG AAG TAC GGG TAT GAT GGG Gln Glu Asp Asn Ser Thr Glu Gly Tyr Trp Lys Tyr Gly Tyr Asp Gly 130 135 140	4320
CAG GAC CAC CTT GAA TTC TGC CCT GAC ACA CTG GAT TGG AGA GCA GCA Gln Asp His Leu Glu Phe Cys Pro Asp Thr Leu Asp Trp Arg Ala Ala 145 150 155	4368
GAA CCC AGG GCC TGG CCC ACC AAG CTG GAG TGG GAA AGG CAC AAG ATT Glu Pro Arg Ala Trp Pro Thr Lys Leu Glu Trp Glu Arg His Lys Ile 160 165 170	4416
CGG GCC AGG CAG AAC AGG GCC TAC CTG GAG AGG GAC TGC CCT GCA CAG Arg Ala Arg Gln Asn Arg Ala Tyr Leu Glu Arg Asp Cys Pro Ala Gln 175 180 185 190	4464
CTG CAG CAG TTG CTG GAG CTG GGG AGA GGT GTT TTG GAC CAA CAA G Leu Gln Gln Leu Leu Glu Leu Gly Arg Gly Val Leu Asp Gln Gln 195 200 205	4510
GTATGGTGGAA AACACACTTC TGCCCCATA CTCTAGTGGC AGAGTGGAGG AGGTTGCAGG	4570
GCACGGAATC CCTGGTTGGA GTTTCAGAGG TGGCTGAGGC TGTGTGCCTC TCCAAATTCT	4630
GGGAAGGGAC TTTCTCAATC CTAGAGTCTC TACCTTATAA TTGAGATGTA TGAGACAGCC	4690
ACAAGTCATG GGTTTAATTT CTTTCTCCA TGCATATGGC TCAAAGGGAA GTGTCTATGG	4750
CCCTTGCTTT TTATTTAACC AATAATCTTT TGTATATTTA TACCTGTTAA AAATTCAGAA	4810
ATGTCAAGGC CGGGCACGGT GGCTCACCCC TGTAATCCCA GCACTTGGG AGGCCGAGGC	4870
GGGTGGTCAC AAGGTCAGGA GTTTGAGACC AGCCTGACCA ACATGGTGAA ACCCGTCTCT	4930
AAAAAAATAC AAAAATTAGC TGGTCACAGT CATGCGCACC TGTAGTCCCA GCTAATTGGA	4990
AGGCTGAGGC AGGAGCATCG CTTGAACCTG GGAAGCGGAA GTTGCAGTGA GCCAAGATCG	5050
CGCCACTGCA CTCCAGCCTA GGCAGCAGAG TGAGACTCCA TCTTAAAAAA AAAAAAAA	5110
AAAAAAAGAG AATTCAAGAGA TCTCAGCTAT CATATGAATA CCAGGACAAA ATATCAAGTG	5170
AGGCCACTTA TCAGAGTAGA AGAATCCTTT AGGTTAAAAG TTTCTTCAT AGAACATAGC	5230

AATAATCACT	GAAGCTACCT	ATCTTACAAG	TCCGCTTCTT	ATAACAATGC	CTCCTAGGTT	5290
GACCCAGGTG	AAACTGACCA	TCTGTATTCA	ATCATTTCA	ATGCACATAA	AGGGCAATT	5350
TATCTATCAG	AACAAAGAAC	ATGGGTAACA	GATATGTATA	TTTACATGTG	AGGAGAACAA	5410
GCTGATCTGA	CTGCTCTCCA	AGTGACACTG	TGTTAGAGTC	CAATCTTAGG	ACACAAAATG	5470
GTGTCTCTCC	TGTAGCTTGT	TTTTTTCTGA	AAAGGGTATT	TCCTTCCTCC	AACCTATAGA	5530
AGGAAGTGAA	AGTTCCAGTC	TTCCCTGGCAA	GGGTAAACAG	ATCCCCCTCTC	CTCATCCTTC	5590
CTCTTTCTG	TCAAG	TG CCT CCT TTG GTG	AAG GTG ACA CAT CAT	GTG ACC		5640
		Val Pro Pro Leu Val	Lys Val Thr His His	Val Thr		
		210	215			
TCT TCA	GTG ACC ACT CTA CGG	TGT CGG GCC	TTG AAC TAC TAC CCC	CAG		5688
Ser Ser Val	Thr Thr Leu Arg	Cys Arg Ala	Leu Asn Tyr Tyr	Pro Gln		
220	225	230				
AAC ATC ACC ATG AAG TGG	CTG AAG GAT AAG CAG CCA	ATG GAT GCC	AAG			5736
Asn Ile Thr Met Lys Trp	Leu Lys Asp Lys Gln	Pro Met Asp Ala	Lys			
235	240	245				
GAG TTC GAA CCT AAA GAC	GTA TTG CCC AAT GGG GAT	GGG ACC TAC CAG				5784
Glu Phe Glu Pro Lys Asp	Val Leu Pro Asn Gly	Asp Gly Thr Tyr	Gln			
250	255	260	265			
GGC TGG ATA ACC TTG GCT	GTA CCC CCT GGG GAA GAG CAG	AGA TAT ACG				5832
Gly Trp Ile Thr Leu Ala	Val Pro Pro Gly	Glu Glu Gln Arg	Tyr Thr			
270	275	280				
TGC CAG GTG GAG CAC CCA	GGC CTG GAT CAG CCC	CTC ATT GTG ATC TGG G				5881
Cys Gln Val Glu His Pro	Gly Leu Asp Gln	Pro Leu Ile Val Ile	Trp			
285	290	295				
GTATGTGACT GATGAGAGCC	AGGAGCTGAG AAAATCTATT	GGGGGTTGAG AGGAGTCCT				5941
GAGGAGGTAA TTATGGCAGT	GAGATGAGGA TCTGCTCTT	GTTAGGGGT GGGCTGAGGG				6001
TGGCAATCAA AGGCTTTAAC	TTGCTTTTC TGTTTAG	AG CCC TCA CCG TCT				6053
		Glu Pro Ser Pro Ser				
		300				
GGC ACC CTA GTC ATT GGA	GTC ATC AGT GGA ATT GCT	GTT TTT GTC GTC				6101
Gly Thr Leu Val Ile	Gly Val Ile Ser Gly	Ile Ala Val Phe Val Val				
305	310	315				
ATC TTG TTC ATT GGA ATT	TTG TTC ATA ATA TTA	AGG AAG AGG CAG GGT				6149
Ile Leu Phe Ile Gly	Ile Leu Phe Ile Ile	Leu Arg Lys Arg Gln Gly				
320	325	330				
TCA A GTGAGTAGGA	ACAAGGGGA AGTCTCTTAG	TACCTCTGCC CCAGGGCACA				6203
Ser						
335						
GTGGGAAGAG	GGGCAGAGGG	GATCTGGCAT CCATGGGAAG	CATTTTCTC	ATTTATATT		6263
TTTGGGGACA	CCAGCAGCTC	CCTGGGAGAC	AGAAAATAAT	GGTTCTCCCC	AGAATGAAAG	6323
TCTCTAATTC	AACAAACATC	TTCAAGAGCAC	CTACTATTTT	GCAAGAGCTG	TTAAGGTAG	6383

TACAGGGGCT TTGAGGTTGA	GAAGTCACTG TGGCTATTCT	CAGAACCCAA ATCTGGTAGG	6443
GAATGAAATT GATAGCAAGT	AAATGTAGTT AAAGAAGACC	CCATGAGGTC CTAAGCAGG	6503
CAGGAAGCAA ATGCTTAGGG	TGTCAAAGGA AAGAATGATC	ACATTCACT GGGGATCAAG	6563
ATAGCCTTCT GGATCTTGAA	GGAGAAGCTG GATTCCATTA	GGTGAGGTTG AAGATGATGG	6623
GAGGTCTACA CAGACGGAGC	AACCATGCCA AGTAGGAGAG	TATAAGGCAT ACTGGGAGAT	6683
TAGAAATAAT TACTGTACCT	TAACCCTGAG TTTGCGTAGC	TATCACTCAC CAATTATGCA	6743
TTTCTACCCC CTGAACATCT	GTGGTGTAGG GAAAAGAGAA	TCAGAAAGAA GCCAGCTCAT	6803
ACAGAGTCCA AGGGTCTTT	GGGATATTGG GTTATGATCA	CTGGGGTGT ATTGAAGGAT	6863
CCTAAGAAAG GAGGACCACG	ATCTCCCTTA TATGGTGAAT	GTGTTGTTAA GAAGTTAGAT	6923
GAGAGGTGAG GAGACCAGTT	AGAAAGCCAA TAAGCATTTC	CAGATGAGAG ATAATGGTTC	6983
TTGAAATCCA ATAGTGCCA	GGTCTAAATT GAGATGGGTG	AATGAGGAAA ATAAGGAAGA	7043
GAGAAGAGGC AAGATGGTGC	CTAGGTTGT GATGCCTCTT	TCCTGGGTCT CTTGTCTCCA	7103
CAG GA GGA GCC ATG GGG CAC TAC GTC TTA GCT GAA CGT GAG			7144
Arg Gly Ala Met Gly His Tyr Val Leu Ala Glu Arg Glu			
340	345		
TGACACCGCAG CCTGCAGACT	CACTGTGGGA AGGAGACAAA	ACTAGAGACT CAAAGAGGGA	7204
GTGCATTTAT GAGCTCTTCA	TGTTTCAGGA GAGAGTTGAA	CCTAAACATA GAAATTGCCT	7264
GACGAACCTCC TTGATTTAG	CCTTCTCTGT TCATTTCCCTC	AAAAAGATTT CCCCATTTAG	7324
GTTTCTGAGT TCCTGCATGC	CGGTGATCCC TAGCTGTGAC	CTCTCCCTG GAACTGTCTC	7384
TCATGAACCT CAAGCTGCAT	CTAGAGGCTT CCTTCATTTTC	CTCCGTCAACC TCAGAGACAT	7444
ACACCTATGT CATTTCATTT	CCTATTTTG GAAGAGGACT	CCTTAAATTG GGGGGACTTA	7504
CATGATTCAT TTTAACATCT	GAGAAAAGCT TTGAACCCCTG	GGACGTGGCT AGTCATAACC	7564
TTACCAGATT TTTACACATG	TATCTATGCA TTTTCTGGAC	CCGTTCAACT TTTCTTTGA	7624
ATCCTCTCTC TGTGTTACCC	AGTAACTCAT CTGTCACCAA	GCCTTGGGA TTCTTCCATC	7684
TGATTGTGAT GTGAGTTGCA	CAGCTATGAA GGCTGTACAC	TGCACGAATG GAAGAGGCAC	7744
CTGTCCCAGA AAAAGCATCA	TGGCTATCTG TGGGTAGTAT	GATGGGTGTT TTTAGCAGGT	7804
AGGAGGCAAA TATCTTGAAA	GGGGTTGTGA AGAGGTGTTT	TTTCTAATTG GCATGAAGGT	7864
GTCATACAGA TTTGCAAAGT	TTAATGGTGC CTTCATTTGG	GATGCTACTC TAGTATTCCA	7924
GACCTGAAGA ATCACAAATAA	TTTCTACCT GGTCTCTCCT	TGTTCTGATA ATGAAAATTA	7984
TGATAAGGAT GATAAAAGCA	CTTACTTCGT GTCCGACTCT	TCTGAGCACC TACTTACATG	8044
CATTACTGCA TGCACCTCTT	ACAATAATTC TATGAGATAG	GTACTATTAT CCCCATTCT	8104
TTTTTAAATG AAGAAAGTGA	AGTAGGCCGG GCACGGTGGC	TCACGCCGT AATCCCAGCA	8164

CTTTGGGAGG CCAAAGCGGG TGGATCACGA GGTCAAGGAGA TCGAGACCAT CCTGGCTAAC	8224
ATGGTGAAAC CCCATCTCTA ATAAAAATAC AAAAAATTAG CTGGGCGTGG TGGCAGACGC	8284
CTGTAGTCCC AGCTACTCGG AAGGCTGAGG CAGGAGAATG GCATGAACCC AGGAGGCAGA	8344
GCTTGCAGTG AGCCGAGTTT GCGCCACTGC ACTCCAGCCT AGGTGACAGA GTGAGACTCC	8404
ATCTCAAAAA AATAAAAATA AAAATAAAA AATGAAAAAA AAAAGAAAGT GAAGTATAGA	8464
GTATCTCATA GTTTGTCAGT GATAGAAACA GGTTCAAAC TCAGTCAATC TGACCGTTG	8524
ATACATCTCA GACACCACTA CATTCACTAG TTTAGATGCC TAGAATAAT AGAGAAGGAA	8584
GGAGATGGCT CTTCTCTTGT CTCATTGTGT TTCTTCTGAG TGAGCTTGAA TCACATGAAG	8644
GGGAACAGCA GAAAACAACC AACTGATCCT CAGCTGTCAT GTTTCCCTTA AAAGTCCCTG	8704
AAGGAAGGTC CTGGAATGTG ACTCCCTTGC TCCTCTGTTG CTCTCTTGG CATTCAATTTC	8764
TTTGGACCCCT ACGCAAGGAC TGTAATTGGT GGGGACAGCT AGTGGCCCTG CTGGGCTTCA	8824
CACACGGTGT CCTCCCTAGG CCAGTGCCTC TGGAGTCAGA ACTCTGGTGG TATTTCCCTC	8884
AATGAAGTGG AGTAAGCTCT CTCATTTGA GATGGTATAA TGGAAGCCAC CAAGTGGCTT	8944
AGAGGATGCC CAGGTCCCTC CATGGAGCCA CTGGGGTTCC GGTGCACATT AAAAAAAA	9004
TCTAACCAAGG ACATTCAAGGA ATTGCTAGAT TCTGGAAAT CAGTCACCA TGTTCAAAG	9064
AGTCTTTTTT TTTTTTTGA GACTCTATTG CCCAGGCTGG AGTGCAATGG CATGATCTCG	9124
GCTCACTGTA ACCTCTGCCT CCCAGGTCA AGCGATTCTC CTGTCTCAGC CTCCCAAGTA	9184
GCTGGGATTA CAGGCGTGCA CCACCATGCC CGGCTAATT TTGTATTTT AGTAGAGACA	9244
GGGTTTCACC ATGTTGGCCA GGCTGGCTC GAACTCTCCT GACCTCGTGA TCCGCCTGCC	9304
TCGGCCTCCC AAAGTGCTGA GATTACAGGT GTGAGCCACC CTGCCAGCC GTCAAAAGAG	9364
TCTTAATATA TATATCCAGA TGGCATGTGT TTACTTTATG TTACTACATG CACTTGGCTG	9424
CATAAAATGTG GTACAAGCAT TCTGTCTTGA AGGGCAGGTG CTTCAGGATA CCATATACAG	9484
CTCAGAAGTT TCTTCTTTAG GCATTAATT TTAGCAAAGA TATCTCATCT CTTCTTTAA	9544
ACCATTTCT TTTTTGTGG TTAGAAAAGT TATGTAGAAA AAAGTAAATG TGATTTACGC	9604
TCATTGTAGA AAAGCTATAA AATGAATACA ATTAAAGCTG TTATTTAATT AGCCAGTGAA	9664
AAACTATTAA CAACTTGTCT ATTACCTGTT AGTATTATTG TTGCATTAAA AATGCATATA	9724
CTTTAATAAA TGTATATTGT ATTGTATACT GCATGATTT ATTGAAGTTC TTGTTCATCT	9784
TGTGTATATA CTTAATCGCT TTGTCATTT GGAGACATT ATTGCTTC TAATTTCTTT	9844
ACATTTGTC TTACGGAATA TTTCAATTCA ACTGTGGTAG CCGAATTAAT CGTGTTCCTT	9904
CACTCTAGGG ACATTGTCGT CTAAGTTGTA AGACATTGGT TATTTACCA GCAAACCATT	9964
CTGAAAGCAT ATGACAAATT ATTTCTCTCT TAATATCTTA CTATACTGAA AGCAGACTGC	10024

TATAAGGCCTT CACTTACTCT TCTACCTCAT AAGGAATATG TTACAATTAA TTTATTAGGT	10084
AAGCATTGT TTTATATTGG TTTTATTCA CCTGGGCTGA GATTCAGA AACACCCCAG	10144
TCTTCACAGT AACACATTTC ACTAACACAT TTACTAAACA TCAGCAACTG TGGCCTGTTA	10204
ATTTTTTAA TAGAAATTT AAGTCCTCAT TTTCTTCGG TGTTTTTAA GCTTAATTT	10264
TCTGGCTTTA TTCATAAATT CTTAAGGTCA ACTACATTG AAAAATCAA GACCTGCATT	10324
TTAAATTCTT ATTCACCTCT GGCAAAACCA TTCACAAACC ATGGTAGTAA AGAGAAGGGT	10384
GACACCTGGT GGCCATAGGT AAATGTACCA CGGTGGTCCG GTGACCAGAG ATGCAGCGCT	10444
GAGGGTTTTC CTGAAGGTAA AGGAATAAAG AATGGGTGGA GGGCGTGCA CTGGAAATCA	10504
CTTGTAGAGA AAAGCCCCTG AAAATTTGAG AAAACAAACA AGAAACTACT TACCAAGCTAT	10564
TTGAATTGCT GGAATCACAG GCCATTGCTG AGCTGCCTGA ACTGGGAACA CAACAGAAGG	10624
AAAACAAACC ACTCTGATAA TCATTGAGTC AAGTACAGCA GGTGATTGAG GACTGCTGAG	10684
AGGTACAGGC CAAAATTCTT ATGTTGTATT ATAATAATGT CATCTTATAA TACTGTCAGT	10744
ATTTTATAAA ACATTCTTC AAAACTCACA CACATTAAA AACAAAACAC TGTCTCTAAA	10804
ATCCCCAAAT TTTTCATAAA C	10825

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 348 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Gly Pro Arg Ala Arg Pro Ala Leu Leu Leu Leu Met Leu Leu Gln
 1 5 10 15

Thr Ala Val Leu Gln Gly Arg Leu Leu Arg Ser His Ser Leu His Tyr
 20 25 30

Leu Phe Met Gly Ala Ser Glu Gln Asp Leu Gly Leu Ser Leu Phe Glu
 35 40 45

Ala Leu Gly Tyr Val Asp Asp Gln Leu Phe Val Phe Tyr Asp Asp Glu
 50 55 60

Ser Arg Arg Val Glu Pro Arg Thr Pro Trp Val Ser Ser Arg Ile Ser
 65 70 75 80

Ser Gln Met Trp Leu Gln Leu Ser Gln Ser Leu Lys Gly Trp Asp His
 85 90 95

Met Phe Thr Val Asp Phe Trp Thr Ile Met Glu Asn His Asn His Ser
 100 105 110

Lys Glu Ser His Thr Leu Gln Val Ile Leu Gly Cys Glu Met Gln Glu

115	120	125
Asp Asn Ser Thr Glu Gly Tyr Trp Lys Tyr Gly Tyr Asp Gly Gln Asp		
130	135	140
His Leu Glu Phe Cys Pro Asp Thr Leu Asp Trp Arg Ala Ala Glu Pro		
145	150	155
160		
Arg Ala Trp Pro Thr Lys Leu Glu Trp Glu Arg His Lys Ile Arg Ala		
165	170	175
Arg Gln Asn Arg Ala Tyr Leu Glu Arg Asp Cys Pro Ala Gln Leu Gln		
180	185	190
Gln Leu Leu Glu Leu Gly Arg Gly Val Leu Asp Gln Gln Val Pro Pro		
195	200	205
Leu Val Lys Val Thr His His Val Thr Ser Ser Val Thr Thr Leu Arg		
210	215	220
Cys Arg Ala Leu Asn Tyr Tyr Pro Gln Asn Ile Thr Met Lys Trp Leu		
225	230	235
240		
Lys Asp Lys Gln Pro Met Asp Ala Lys Glu Phe Glu Pro Lys Asp Val		
245	250	255
Leu Pro Asn Gly Asp Gly Thr Tyr Gln Gly Trp Ile Thr Leu Ala Val		
260	265	270
Pro Pro Gly Glu Glu Gln Arg Tyr Thr Cys Gln Val Glu His Pro Gly		
275	280	285
Leu Asp Gln Pro Leu Ile Val Ile Trp Glu Pro Ser Pro Ser Gly Thr		
290	295	300
Leu Val Ile Gly Val Ile Ser Gly Ile Ala Val Phe Val Val Ile Leu		
305	310	315
320		
Phe Ile Gly Ile Leu Phe Ile Ile Leu Arg Lys Arg Gln Gly Ser Arg		
325	330	335
Gly Ala Met Gly His Tyr Val Leu Ala Glu Arg Glu		
340	345	

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10825 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)

- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: join(361..436, 3762..4025, 4235..4510, 5606..5881, 6040..6153, 7107..7147)
 - (D) OTHER INFORMATION: /product= "Hereditary Hemochromatosis (HH) protein containing both the 24d1

and 24d2 mutations"
 /note= "Hereditary Hemochromatosis (HH)
 gene containing a combination of both
 24d1 and 24d2 alleles"

(ix) FEATURE:

(A) NAME/KEY: -
 (B) LOCATION: 140..7319
 (D) OTHER INFORMATION: /note= "start and stop positions for
 cDNA containing a combination of both
 24d1 and 24d2 alleles
 (SEQ ID NO:12)"

(ix) FEATURE:

(A) NAME/KEY: -
 (B) LOCATION: 3852..3891
 (D) OTHER INFORMATION: /note= "start and stop positions for
 genomic sequence surrounding variant
 for 24d2(G) allele (SEQ ID NO:42)"

(ix) FEATURE:

(A) NAME/KEY: -
 (B) LOCATION: 5507..6023
 (D) OTHER INFORMATION: /note= "start and stop positions for
 genomic sequence surrounding variant
 for 24d1(A) allele (SEQ ID NO:21)"

(ix) FEATURE:

(A) NAME/KEY: allele
 (B) LOCATION: replace(3872, "g")
 (D) OTHER INFORMATION: /phenotype= "Hereditary Hemochromatosis
 (HH)"
 /label= 24d2

(ix) FEATURE:

(A) NAME/KEY: allele
 (B) LOCATION: replace(5834, "a")
 (D) OTHER INFORMATION: /phenotype= "Hereditary Hemochromatosis
 (HH)"
 /label= 24d1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

TCTAAGGTTG AGATAAAATT TTTAAATGTA TGATTGAATT TTGAAAATCA TAAATATTAA	60
AATATCTAAA GTTCAGATCA GAACATTGCG AAGCTACTTT CCCCAATCAA CAACACCCCT	120
TCAGGATTAA AAAACCAAGG GGGACACTGG ATCACCTAGT GTTTCACAAG CAGGTACCTT	180
CTGCTGTAGG AGAGAGAGAA CTAAAGTTCT GAAAGACCTG TTGCTTTCA CCAGGAAGTT	240
TTACTGGGCA TCTCCTGAGC CTAGGCAATA GCTGTAGGGT GACTTCTGGA GCCATCCCCG	300
TTTCCCCGCC CCCCAAAAGA AGCGGAGATT TAACGGGGAC GTGCGGCCAG AGCTGGGGAA	360
ATG GGC CCG CGA GCC AGG CCG GCG CTT CTC CTC CTG ATG CTT TTG CAG Met Gly Pro Arg Ala Arg Pro Ala Leu Leu Leu Leu Met Leu Leu Gln	408
1 5 10 15	
ACC GCG GTC CTG CAG GGG CGC TTG CTG C GTGAGTCCGA GGGCTGCGGG Thr Ala Val Leu Gln Gly Arg Leu Leu	456

20

25

CGAACTAGGG GCGCGGCGGG GGTGGAAAAA TCGAAACTAG CTTTTCTTT GCGCTTGGGA	516
GTGGCTAAC TTTGGAGGAC CTGCTCAACC CTATCCGCAA GCCCCTCTCC CTACTTTCTG	576
CGTCCAGACC CCGTGAGGGA GTGCCTACCA CTGAACGTCA GATAGGGGTC CCTCGCCCCA	636
GGACCTGCC CTCCTCCGG CTGTCCCGC TCTGCGGAGT GACTTTGGA ACCGCCACT	696
CCCTCCCCC AACTAGAATG CTTTTAAATA AATCTCGTAG TTCCTCACTT GAGCTGAGCT	756
AAGCCTGGGG CTCCTTGAAC CTGGAACTCG GGTTTATTTC CAATGTCAGC TGTGCAGTTT	816
TTTCCCCAGT CATCTCCAAA CAGGAAGTTC TTCCCTGAGT GCTTGCCGAG AAGGCTGAGC	876
AAACCCACAG CAGGATCCGC ACGGGGTTTC CACCTCAGAA CGAATGCGTT GGGCGGTGGG	936
GGCGCGAAAG AGTGGCGTTG GGGATCTGAA TTCTTCACCA TTCCACCCAC TTTTGGTGAG	996
ACCTGGGTG GAGGTCTCTA GGGTGGGAGG CTCCTGAGAG AGGCCTACCT CGGGCCTTTC	1056
CCCACTCTTG GCAATTGTTTC TTTGCCTGG AAAATTAAGT ATATGTTAGT TTTGAACGTT	1116
TGAACTGAAC AATTCTCTTT TCGGCTAGGC TTTATTGATT TGCAATGTGC TGTGTAATT	1176
AGAGGCCTCT CTACAAAGTA CTGATAATGA ACATGTAAGC AATGCACTCA CTTCTAAGTT	1236
ACATTCAAT CTGATCTTAT TTGATTTCA CTAGGCATAG GGAGGTAGGA GCTAATAATA	1296
CGTTTATTTC ACTAGAAGTT AACTGGAATT CAGATTATAT AACTCTTTTC AGGTTACAAA	1356
GAACATAAAAT AATCTGGTTT TCTGATGTTA TTTCAAGTAC TACAGCTGCT TCTAATCTTA	1416
GTTGACAGTG ATTTGCCCT GTAGTGTAGC ACAGTGTCT GTGGGTACCA CGCCGGCCTC	1476
AGCACAGCAC TTTGAGTTTT GGTACTACGT GTATCCACAT TTTACACATG ACAAGAATGA	1536
GGCATGGCAC GGCCTGCTTC CTGGCAAATT TATTCAATGG TACACTGGC TTTGGTGGCA	1596
GAGCTCATGT CTCCACTTCA TAGCTATGAT TCTTAAACAT CACACTGCAT TAGAGGTTGA	1656
ATAATAAAAT TTCATGTTGA GCAGAAATAT TCATTGTTA CAAGTGTAAA TGAGTCCCAG	1716
CCATGTGTTG CACTGTTCAA GCCCCAAGGG AGAGAGCAGG GAAACAAGTC TTTACCCCTT	1776
GATATTTGC ATTCTAGTGG GAGAGATGAC AATAAGCAA TGAGCAGAAA GATATACAAC	1836
ATCAGGAAAT CATGGGTGTT GTGAGAAGCA GAGAAGTCAG GGCAAGTCAC TCTGGGCTG	1896
ACACTTGAGC AGAGACATGA AGGAAATAAG AATGATATTG ACTGGGAGCA GTATTTCCCA	1956
GGCAAACATGA GTGGGCCTGG CAAGTTGGAT TAAAAAGCGG GTTTCTCAG CACTACTCAT	2016
GTGTGTGTGT GTGGGGGGGG GGGCGGGCGT GGGGGTGGGA AGGGGGACTA CCATCTGCAT	2076
GTAGGATGTC TAGCAGTATC CTGTCCTCCC TACTCACTAG GTGCTAGGAG CACTCCCCA	2136
GTCTTGACAA CCAAAATGT CTCTAAACTT TGCCACATGT CACCTAGTAG ACAAAACTCCT	2196
GGTTAAGAAG CTCGGGTTGA AAAAAATAAA CAAGTAGTGC TGGGGAGTAG AGGCCAAGAA	2256

GTAGGTAATG GGCTCAGAAG AGGAGCCACA AACAAAGGTTG TGCAGGCAGC TGTAGGCTGT	2316
GGTGTGAATT CTAGCCAAGG AGTAACAGTG ATCTGTCACA GGCTTTAAA AGATTGCTCT	2376
GGCTGCTATG TGGAAAGCAG AATGAAGGGA GCAACAGTAA AAGCAGGGAG CCCAGCCAGG	2436
AAGCTGTTAC ACAGTCCAGG CAAGAGGTAG TGGAGTGGC TGGGTGGAA CAGAAAAGGG	2496
AGTGACAAAC CATTGTCTCC TGAATATATT CTGAAGGAAG TTGCTGAAGG ATTCTATGTT	2556
GTGTGAGAGA AAGAGAAGAA TTGGCTGGGT GTAGTAGCTC ATGCCAAGGA GGAGGCCAAG	2616
GAGAGCAGAT TCCTGAGCTC AGGAGTTCAA GACCAGCCTG GGCAACACAG CAAAACCCCT	2676
TCTCTACAAA AAATACAAAA ATTAGCTGGG TGTGGTGGCA TGCACCTGTG ATCCTAGCTA	2736
CTCGGGAGGC TGAGGTGGAG GGTATTGCTT GAGCCCAGGA AGTTGAGGCT GCAGTGAGCC	2796
ATGACTGTGC CACTGTACTT CAGCCTAGGT GACAGAGCAA GACCCTGTCT CCCCTGACCC	2856
CCTGAAAAAG AGAAGAGTTA AAGTTGACTT TGTTCTTTAT TTTAATTAA TTGGCCTGAG	2916
CAGTGGGGTA ATTGGCAATG CCATTTCTGA GATGGTGAAG GCAGAGGAAA GAGCAGTTG	2976
GGGTAAATCA AGGATCTGCA TTTGGACAT GTTAAGTTG AGATTCCAGT CAGGCTTCCA	3036
AGTGGTGAGG CCACATAGGC AGTCAGTGT AAGAATTCA GACCAAGGCT GGGCACGGTG	3096
GCTCACTTCT GTAATCCCAG CACTTTGGTG GCTGAGGCAG GTAGATCATT TGAGGTCAGG	3156
AGTTTGAGAC AAGCTTGGCC AACATGGTGA AACCCATGT CTACTAAAAA TACAAAAATT	3216
AGCCTGGTGT GGTGGCGCAC GCCTATAGTC CCAGGTTTTC AGGAGGCTTA GGTAGGAGAA	3276
TCCCTTGAAC CCAGGAGGTG CAGGTTGCAG TGAGCTGAGA TTGTGCCACT GCACTCCAGC	3336
CTGGGTGATA GAGTGAGACT CTGCTCAAA AAAAAAAA AAAAAAAA AAAAAACTGA	3396
AGGAATTATT CCTCAGGATT TGGGTCTAAT TTGCCCTGAG CACCAACTCC TGAGTTCAAC	3456
TACCATGGCT AGACACACCT TAACATTTC TAGAATCCAC CAGCTTAGT GGAGTCTGTC	3516
TAATCATGAG TATTGGAATA GGATCTGGGG GCAGTGAGGG GGTGGCAGCC ACGTGTGGCA	3576
GAGAAAAGCA CACAAGGAAA GAGCACCCAG GACTGTCATA TGGAAGAAAG ACAGGACTGC	3636
AACTCACCCCT TCACAAAATG AGGACCAGAC ACAGCTGATG GTATGAGTTG ATGCAGGTGT	3696
GTGGAGCCTC AACATCCTGC TCCCCCTCTA CTACACATGG TTAAGGCCTG TTGCTCTGTC	3756
TCCAG GT TCA CAC TCT CTG CAC TAC CTC TTC ATG GGT GCC TCA GAG	3802
Arg Ser His Ser Leu His Tyr Leu Phe Met Gly Ala Ser Glu	
30 35	
CAG GAC CTT GGT CTT TCC TTG TTT GAA GCT TTG GGC TAC GTG GAT GAC	3850
Gln Asp Leu Gly Leu Ser Leu Phe Glu Ala Leu Gly Tyr Val Asp Asp	
40 45 50 55	
CAG CTG TTC GTG TTC TAT GAT GAG AGT CGC CGT GTG GAG CCC CGA	3898
Gln Leu Phe Val Phe Tyr Asp Asp Glu Ser Arg Arg Val Glu Pro Arg	
60 65 70	

ACT CCA TGG GTT TCC AGT AGA ATT TCA AGC CAG ATG TGG CTG CAG CTG Thr Pro Trp Val Ser Ser Arg Ile Ser Ser Gln Met Trp Leu Gln Leu 75 80 85	3946
AGT CAG AGT CTG AAA GGG TGG GAT CAC ATG TTC ACT GTT GAC TTC TGG Ser Gln Ser Leu Lys Gly Trp Asp His Met Phe Thr Val Asp Phe Trp 90 95 100	3994
ACT ATT ATG GAA AAT CAC AAC CAC AGC AAG G GTATGTGGAG AGGGGGCCTC Thr Ile Met Glu Asn His Asn His Ser Lys 105 110	4045
ACCTTCCTGA GGTTGTCAGA GCTTTTCATC TTTTCATGCA TCTTGAAGGA AACAGCTGGA	4105
AGTCTGAGGT CTTGTGGGAG CAGGGAAAGAG GGAAGGAATT TGCTTCCTGA GATCATTGG	4165
TCCTTGGGAA TGGTGGAAAT AGGGACCTAT TCCTTGGTT GCAGTTAACAA AGGCTGGGA	4225
TTTTTCCAG AG TCC CAC ACC CTG CAG GTC ATC CTG GGC TGT GAA ATG Glu Ser His Thr Leu Gln Val Ile Leu Gly Cys Glu Met 115 120 125	4272
CAA GAA GAC AAC AGT ACC GAG GGC TAC TGG AAG TAC GGG TAT GAT GGG Gln Glu Asp Asn Ser Thr Glu Gly Tyr Trp Lys Tyr Gly Tyr Asp Gly 130 135 140	4320
CAG GAC CAC CTT GAA TTC TGC CCT GAC ACA CTG GAT TGG AGA GCA GCA Gln Asp His Leu Glu Phe Cys Pro Asp Thr Leu Asp Trp Arg Ala Ala 145 150 155	4368
GAA CCC AGG GCC TGG CCC ACC AAG CTG GAG TGG GAA AGG CAC AAG ATT Glu Pro Arg Ala Trp Pro Thr Lys Leu Glu Trp Glu Arg His Lys Ile 160 165 170	4416
CGG GCC AGG CAG AAC AGG GCC TAC CTG GAG AGG GAC TGC CCT GCA CAG Arg Ala Arg Gln Asn Arg Ala Tyr Leu Glu Arg Asp Cys Pro Ala Gln 175 180 185 190	4464
CTG CAG CAG TTG CTG GAG CTG GGG AGA GGT GTT TTG GAC CAA CAA G Leu Gln Leu Leu Glu Leu Gly Arg Gly Val Leu Asp Gln Gln 195 200 205	4510
GTATGGTGGAA AACACACTTC TGCCCCTATA CTCTAGTGGC AGAGTGGAGG AGGTTGCAGG	4570
GCACGGAATC CCTGGTTGGA GTTTCAGAGG TGGCTGAGGC TGTGTGCCCTC TCCAAATTCT	4630
GGGAAGGGAC TTTCTCAATC CTAGAGTCTC TACCTTATAA TTGAGATGTA TGAGACAGCC	4690
ACAAGTCATG GGTTTAATTCTT CTTTCTCCA TGCATATGGC TCAAAGGGAA GTGTCTATGG	4750
CCCTTGCTTT TTATTTAACCA AATAATCTTT TGTATATTTA TACCTGTTAA AAATTCAGAA	4810
ATGTCAAGGC CGGGCACGGT GGCTCACCCC TGTAATCCCA GCACCTTGGG AGGCCGAGGC	4870
GGGTGGTCAC AAGGTCAGGA GTTGAGACC AGCCTGACCA ACATGGTGAA ACCCGTCTCT	4930
AAAAAAATAC AAAAATTAGC TGGTCACAGT CATGCGCACC TGTAGTCCCA GCTAATTGGA	4990
AGGCTGAGGC AGGAGCATCG CTTGAACCTG GGAAGCGGAA GTTGCAGTGA GCCAAGATCG	5050
CGCCACTGCA CTCCAGCCTA GGCAGCAGAG TGAGACTCCA TCTTAAAAAA AAAAAAAA	5110

TTTGGGGACA CCAGCAGCTC CCTGGGAGAC AGAAAATAAT GGTTCTCCCC AGAATGAAAG	6323
TCTCTAATTC AACAAACATC TTCAGAGCAC CTACTATTT GCAAGAGCTG TTTAAGGTAG	6383
TACAGGGCT TTGAGGTTGA GAAGTCACTG TGGCTATTCT CAGAACCAA ATCTGGTAGG	6443
GAATGAAATT GATAGCAAGT AAATGTAGTT AAAGAAGACC CCATGAGGTC CTAAAGCAGG	6503
CAGGAAGCAA ATGCTTAGGG TGTCAAAGGA AAGAATGATC ACATTCACTGGGATCAAG	6563
ATAGCCTTCT GGATCTTGAA GGAGAAGCTG GATTCCATTA GGTGAGGTTG AAGATGATGG	6623
GAGGTCTACA CAGACGGAGC AACCATGCCA AGTAGGAGAG TATAAGGCAT ACTGGGAGAT	6683
TAGAAATAAT TACTGTACCT TAACCCTGAG TTTGCGTAGC TATCACTCAC CAATTATGCA	6743
TTTCTACCCC CTGAACATCT GTGGTGTAGG GAAAAGAGAA TCAGAAAGAA GCCAGCTCAT	6803
ACAGAGTCCA AGGGCTTTT GGGATATTGG GTTATGATCA CTGGGGTGTC ATTGAAGGAT	6863
CCTAAGAAAG GAGGACCAAG ATCTCCCTTA TATGGTGAAT GTGTTGTTAA GAAGTTAGAT	6923
GAGAGGTGAG GAGACCAGTT AGAAAGCAA TAAGCATTTC CAGATGAGAG ATAATGGTTC	6983
TTGAAATCCA ATAGTGCCCA GGTCTAAATT GAGATGGGTG AATGAGGAAA ATAAGGAAGA	7043
GAGAAGAGGC AAGATGGTGC CTAGGTTGT GATGCCTCTT TCCTGGGTCT CTTGTCTCCA	7103
CAG GA GGA GCC ATG GGG CAC TAC GTC TTA GCT GAA CGT GAG	7144
Arg Gly Ala Met Gly His Tyr Val Leu Ala Glu Arg Glu	
340	345
TGACACCGCAG CCTGCAGACT CACTGTGGGA AGGAGACAAA ACTAGAGACT CAAAGAGGGAA	7204
GTGCATTAT GAGCTCTCA TGTTTCAGGA GAGAGTTGAA CCTAAACATA GAAATTGCCT	7264
GACGAACCTCC TTGATTTAG CCTTCTCTGT TCATTTCTC AAAAAGATT CCCCATTAG	7324
GTTCCTGAGT TCCTGCATGC CGGTGATCCC TAGCTGTGAC CTCTCCCTG GAACTGTCTC	7384
TCATGAACCT CAAGCTGCAT CTAGAGGCTT CCTTCATTTC CTCCGTACCC TCAGAGACAT	7444
ACACCTATGT CATTTCATTT CCTATTTTG GAAGAGGACT CCTTAAATTG GGGGGACTTA	7504
CATGATTCAT TTTAACATCT GAGAAAAGCT TTGAACCTG GGACGTGGCT AGTCATAACC	7564
TTACCCAGATT TTTACACATG TATCTATGCA TTTCTGGAC CCGTTCAACT TTTCTTTGA	7624
ATCCTCTCTC TGTGTTACCC AGTAACTCAT CTGTCACCAA GCCTTGGGA TTCTTCCATC	7684
TGATTGTGAT GTGAGTTGCA CAGCTATGAA GGCTGTACAC TGCACGAATG GAAGAGGCAC	7744
CTGTCCCAGA AAAAGCATCA TGGCTATCTG TGGGTAGTAT GATGGGTGTT TTTAGCAGGT	7804
AGGAGGCAA TATCTTGAAA GGGGTTGTGA AGAGGTGTTT TTTCTAATTG GCATGAAGGT	7864
GTCATACAGA TTTGCAAAGT TTAATGGTGC CTTCAATTGG GATGCTACTC TAGTATTCCA	7924
GACCTGAAGA ATCACAAATAA TTTTCTACCT GGTCTCTCCT TGTTCTGATA ATGAAAATTAA	7984
TGATAAGGAT GATAAAAGCA CTTACTTCGT GTCCGACTCT TCTGAGCACC TACTTACATG	8044

CATTACTGCA TGCACTTCTT ACAATAATTC TATGAGATAG GTACTATTAT CCCCATTCT	8104
TTTTTAAATG AAGAAAGTGA AGTAGGCCGG GCACGGTGGC TCACGCCTGT AATCCCAGCA	8164
CTTTGGGAGG CCAAAGCGGG TGGATCACGA GGTCAAGGAGA TCGAGACCAT CCTGGCTAAC	8224
ATGGTGAAAC CCCATCTCTA ATAAAAATAC AAAAAATTAG CTGGCGTGG TGGCAGACGC	8284
CTGTAGTCCC AGCTACTCGG AAGGCTGAGG CAGGAGAATG GCATGAACCC AGGAGGCAGA	8344
GCTTGCAGTG AGCCGAGTTT GCGCCACTGC ACTCCAGCCT AGGTGACAGA GTGAGACTCC	8404
ATCTCAAAAA AATAAAAATA AAAATAAAA AATGAAAAAA AAAAGAAAGT GAAGTATAGA	8464
GTATCTCATA GTTTGTCAGT GATAGAAACA GGTTCAAAC TCAGTCAATC TGACCGTTG	8524
ATACATCTCA GACACCACTA CATTCACTAG TTTAGATGCC TAGAATAAAT AGAGAAGGAA	8584
GGAGATGGCT CTTCTCTGT CTCATTGTGT TTCTTCTGAG TGAGCTTGAA TCACATGAAG	8644
GGGAACAGCA GAAAACAACC AACTGATCCT CAGCTGTCA GTTCCCTTA AAAGTCCCTG	8704
AAGGAAGGTC CTGGAATGTG ACTCCCTTGC TCCTCTGTTG CTCTCTTGG CATTCAATTTC	8764
TTTGGACCT ACAGCAAGGAC TGTAATTGGT GGGGACAGCT AGTGGCCCTG CTGGGCTTCA	8824
CACACGGTGT CCTCCCTAGG CCAGTGCCTC TGGAGTCAGA ACTCTGGTGG TATTTCCCTC	8884
AATGAAGTGG AGTAAGCTCT CTCATTTGA GATGGTATAA TGGAAAGCCAC CAAGTGGCTT	8944
AGAGGATGCC CAGGCCCTTC CATGGAGCCA CTGGGGTTCC GGTGCACATT AAAAAAAA	9004
TCTAACCAAGG ACATTCAAGGA ATTGCTAGAT TCTGGGAAAT CAGTTCACCA TGTCAAAAG	9064
AGTCTTTTTT TTTTTTTTGA GACTCTATTG CCCAGGCTGG AGTGAATGG CATGATCTCG	9124
GCTCACTGTA ACCTCTGCCT CCCAGGTTCA AGCGATTCTC CTGTCTCAGC CTCCCAAGTA	9184
GCTGGGATTA CAGGCGTGCA CCACCATGCC CGGCTAATT TTGTATTTT AGTAGAGACA	9244
GGGTTTCACC ATGTTGGCCA GGCTGGTCTC GAACTCTCCT GACCTCGTGA TCCGCCTGCC	9304
TCGGCCTCCC AAAGTGTGA GATTACAGGT GTGAGCCACC CTGCCAGCC GTCAAAAGAG	9364
TCTTAATATA TATATCCAGA TGGCATGTGT TTACTTTATG TTACTACATG CACTTGGCTG	9424
CATAAAATGTG GTACAAGCAT TCTGTCTTGA AGGGCAGGTG CTTCAGGATA CCATATACAG	9484
CTCAGAAGTT TCTTCTTTAG GCATTAATT TTAGCAAAGA TATCTCATCT CTTCTTTAA	9544
ACCATTTCT TTTTTTGTGG TTAGAAAAGT TATGTAGAAA AAAGTAAATG TGATTTACGC	9604
TCATTGTAGA AAAGCTATAA AATGAATACA ATTAAAGCTG TTATTTAATT AGCCAGTGAA	9664
AAACTATTAA CAACTTGTCT ATTACCTGTT AGTATTATTG TTGCATTAAA AATGCATATA	9724
CTTTAATAAA TGTATATTGT ATTGTATACT GCATGATTT ATTGAAGTTC TTGTTCATCT	9784
TGTGTATATA CTTAACCGCT TTGTCATTTC GGAGACATTT ATTTGCTTC TAATTTCTTT	9844
ACATTTGTC TTACGGAATA TTTTCATTCA ACTGTGGTAG CCGAATTAAT CGTGTTC	9904

CACTCTAGGG ACATTGTCGT CTAAGTTGTA AGACATTGGT TATTTTACCA GCAAACCATT	9964
CTGAAAGCAT ATGACAAATT ATTTCTCTCT TAATATCTTA CTATACTGAA AGCAGACTGC	10024
TATAAGGCTT CACTTACTCT TCTACCTCAT AAGGAATATG TTACAATTAA TTTATTAGGT	10084
AAGCATTGTT TTTATATTGG TTTTATTCA CCTGGGCTGA GATTTCAGA AACACCCAG	10144
TCTTCACAGT AACACATTTC ACTAACACAT TTACTAAACA TCAGCAACTG TGGCCTGTTA	10204
ATTTTTTAA TAGAAATTT AAGTCCTCAT TTTCTTCGG TGTTTTTAA GCTTAATTAA	10264
TCTGGCTTTA TTCATAAATT CTTAAGGTCA ACTACATTG AAAATCAA GACCTGCATT	10324
TTAAATTCTT ATTCACCTCT GGCAAAACCA TTCACAAACC ATGGTAGTAA AGAGAAGGGT	10384
GACACCTGGT GGCCATAGGT AAATGTACCA CGGTGGTCCG GTGACCAGAG ATGCAGCGCT	10444
GAGGGTTTC CTGAAGGTAA AGGAATAAAG AATGGGTGGA GGGGCGTGCA CTGGAAATCA	10504
CTTGTAGAGA AAAGCCCTG AAAATTGAG AAAACAAACA AGAAACTACT TACCAAGCTAT	10564
TTGAATTGCT GGAATCACAG GCCATTGCTG AGCTGCCTGA ACTGGGAACA CAACAGAAGG	10624
AAAACAAACC ACTCTGATAA TCATTGAGTC AAGTACAGCA GGTGATTGAG GACTGCTGAG	10684
AGGTACAGGC CAAAATTCTT ATGTTGTATT ATAATAATGT CATCTTATAA TACTGTCAGT	10744
ATTTTATAAA ACATTCTCA CAAACTCACA CACATTAAA AACAAACAC TGTCTCTAAA	10804
ATCCCCAAAT TTTTCATAAA C	10825

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 348 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met	Gly	Pro	Arg	Ala	Arg	Pro	Ala	Leu	Leu	Leu	Leu	Met	Leu	Leu	Gln
1								5				10			15
Thr	Ala	Val	Leu	Gln	Gly	Arg	Leu	Leu	Arg	Ser	His	Ser	Leu	His	Tyr
								20				25			30
Leu	Phe	Met	Gly	Ala	Ser	Glu	Gln	Asp	Leu	Gly	Leu	Ser	Leu	Phe	Glu
								35				40			45
Ala	Leu	Gly	Tyr	Val	Asp	Asp	Gln	Leu	Phe	Val	Phe	Tyr	Asp	Asp	Glu
								50				55			60
Ser	Arg	Arg	Val	Glu	Pro	Arg	Thr	Pro	Trp	Val	Ser	Ser	Arg	Ile	Ser
								65				70			80
Ser	Gln	Met	Trp	Leu	Gln	Leu	Ser	Gln	Ser	Leu	Lys	Gly	Trp	Asp	His
								85				90			95

Met Phe Thr Val Asp Phe Trp Thr Ile Met Glu Asn His Asn His Ser
 100 105 110

Lys Glu Ser His Thr Leu Gln Val Ile Leu Gly Cys Glu Met Gln Glu
 115 120 125

Asp Asn Ser Thr Glu Gly Tyr Trp Lys Tyr Gly Tyr Asp Gly Gln Asp
 130 135 140

His Leu Glu Phe Cys Pro Asp Thr Leu Asp Trp Arg Ala Ala Glu Pro
 145 150 155 160

Arg Ala Trp Pro Thr Lys Leu Glu Trp Glu Arg His Lys Ile Arg Ala
 165 170 175

Arg Gln Asn Arg Ala Tyr Leu Glu Arg Asp Cys Pro Ala Gln Leu Gln
 180 185 190

Gln Leu Leu Glu Leu Gly Arg Gly Val Leu Asp Gln Gln Val Pro Pro
 195 200 205

Leu Val Lys Val Thr His His Val Thr Ser Ser Val Thr Thr Leu Arg
 210 215 220

Cys Arg Ala Leu Asn Tyr Tyr Pro Gln Asn Ile Thr Met Lys Trp Leu
 225 230 235 240

Lys Asp Lys Gln Pro Met Asp Ala Lys Glu Phe Glu Pro Lys Asp Val
 245 250 255

Leu Pro Asn Gly Asp Gly Thr Tyr Gln Gly Trp Ile Thr Leu Ala Val
 260 265 270

Pro Pro Gly Glu Glu Gln Arg Tyr Thr Tyr Gln Val Glu His Pro Gly
 275 280 285

Leu Asp Gln Pro Leu Ile Val Ile Trp Glu Pro Ser Pro Ser Gly Thr
 290 295 300

Leu Val Ile Gly Val Ile Ser Gly Ile Ala Val Phe Val Val Ile Leu
 305 310 315 320

Phe Ile Gly Ile Leu Phe Ile Ile Leu Arg Lys Arg Gln Gly Ser Arg
 325 330 335

Gly Ala Met Gly His Tyr Val Leu Ala Glu Arg Glu
 340 345

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1440 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: CDS

(B) LOCATION: 222..1268

(ix) FEATURE:

(A) NAME/KEY: allele
 (B) LOCATION: replace(408, "c")
 (D) OTHER INFORMATION: /phenotype= "normal or wild-type
 (unaffected)"
 /label= 24d2

(ix) FEATURE:

(A) NAME/KEY: allele
 (B) LOCATION: replace(414, "a")
 (D) OTHER INFORMATION: /phenotype= "normal or wild-type
 (unaffected)"
 /label= 24d7

(ix) FEATURE:

(A) NAME/KEY: allele
 (B) LOCATION: replace(1066, "g")
 (D) OTHER INFORMATION: /phenotype= "normal or wild-type
 (unaffected)"
 /label= 24d1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGGGACACTG GATCACCTAG TGTTTCACAA GCAGGTACCT TCTGCTGTAG GAGAGAGAGA	60
ACTAAAGTTC TGAAAGACCT GTTGCTTTTC ACCAGGAAGT TTTACTGGGC ATCTCCTGAG	120
CCTAGGCAAT AGCTGTAGGG TGACTTCTGG AGCCATCCCC GTTTCCCCGC CCCCCAAAAG	180
AAGCGGAGAT TTAACGGGGA CGTGCAGCCA GAGCTGGGGA A ATG GGC CCG CGA Met Gly Pro Arg	233
1	
GCC AGG CCG GCG CTT CTC CTC CTG ATG CTT TTG CAG ACC GCG GTC CTG Ala Arg Pro Ala Leu Leu Leu Met Leu Leu Gln Thr Ala Val Leu	281
5 10 15 20	
CAG GGG CGC TTG CTG CGT TCA CAC TCT CTG CAC TAC CTC TTC ATG GGT Gln Gly Arg Leu Leu Arg Ser His Ser Leu His Tyr Leu Phe Met Gly	329
25 30 35	
GCC TCA GAG CAG GAC CTT GGT CTT TCC TTG TTT GAA GCT TTG GGC TAC Ala Ser Glu Gln Asp Leu Gly Leu Ser Leu Phe Glu Ala Leu Gly Tyr	377
40 45 50	
GTG GAT GAC CAG CTG TTC GTG TTC TAT GAT CAT GAG AGT CGC CGT GTG Val Asp Asp Gln Leu Phe Val Phe Tyr Asp His Glu Ser Arg Arg Val	425
55 60 65	
GAG CCC CGA ACT CCA TGG GTT TCC AGT AGA ATT TCA AGC CAG ATG TGG Glu Pro Arg Thr Pro Trp Val Ser Ser Arg Ile Ser Ser Gln Met Trp	473
70 75 80	
CTG CAG CTG AGT CAG AGT CTG AAA GGG TGG GAT CAC ATG TTC ACT GTT Leu Gln Leu Ser Gln Ser Leu Lys Gly Trp Asp His Met Phe Thr Val	521
85 90 95 100	
GAC TTC TGG ACT ATT ATG GAA AAT CAC AAC CAC AGC AAG GAG TCC CAC	569

Asp Phe Trp Thr Ile Met Glu Asn His Asn His Ser Lys Glu Ser His			
105	110	115	
ACC CTG CAG GTC ATC CTG GGC TGT GAA ATG CAA GAA GAC AAC AGT ACC		617	
Thr Leu Gln Val Ile Leu Gly Cys Glu Met Gln Glu Asp Asn Ser Thr			
120	125	130	
GAG GGC TAC TGG AAG TAC GGG TAT GAT GGG CAG GAC CAC CTT GAA TTC		665	
Glu Gly Tyr Trp Lys Tyr Gly Tyr Asp Gly Gln Asp His Leu Glu Phe			
135	140	145	
TGC CCT GAC ACA CTG GAT TGG AGA GCA GCA GAA CCC AGG GCC TGG CCC		713	
Cys Pro Asp Thr Leu Asp Trp Arg Ala Ala Glu Pro Arg Ala Trp Pro			
150	155	160	
ACC AAG CTG GAG TGG GAA AGG CAC AAG ATT CGG GCC AGG CAG AAC AGG		761	
Thr Lys Leu Glu Trp Glu Arg His Lys Ile Arg Ala Arg Gln Asn Arg			
165	170	175	180
GCC TAC CTG GAG AGG GAC TGC CCT GCA CAG CTG CAG CAG TTG CTG GAG		809	
Ala Tyr Leu Glu Arg Asp Cys Pro Ala Gln Leu Gln Gln Leu Leu Glu			
185	190	195	
CTG GGG AGA GGT GTT TTG GAC CAA CAA GTG CCT CCT TTG GTG AAG GTG		857	
Leu Gly Arg Gly Val Leu Asp Gln Gln Val Pro Pro Leu Val Lys Val			
200	205	210	
ACA CAT CAT GTG ACC TCT TCA GTG ACC ACT CTA CGG TGT CGG GCC TTG		905	
Thr His His Val Thr Ser Ser Val Thr Thr Leu Arg Cys Arg Ala Leu			
215	220	225	
AAC TAC TAC CCC CAG AAC ATC ACC ATG AAG TGG CTG AAG GAT AAG CAG		953	
Asn Tyr Tyr Pro Gln Asn Ile Thr Met Lys Trp Leu Lys Asp Lys Gln			
230	235	240	
CCA ATG GAT GCC AAG GAG TTC GAA CCT AAA GAC GTA TTG CCC AAT GGG		1001	
Pro Met Asp Ala Lys Glu Phe Glu Pro Lys Asp Val Leu Pro Asn Gly			
245	250	255	260
GAT GGG ACC TAC CAG GGC TGG ATA ACC TTG GCT GTA CCC CCT GGG GAA		1049	
Asp Gly Thr Tyr Gln Gly Trp Ile Thr Leu Ala Val Pro Pro Gly Glu			
265	270	275	
GAG CAG AGA TAT ACG TGC CAG GTG GAG CAC CCA GGC CTG GAT CAG CCC		1097	
Glu Gln Arg Tyr Thr Cys Gln Val Glu His Pro Gly Leu Asp Gln Pro			
280	285	290	
CTC ATT GTG ATC TGG GAG CCC TCA CCG TCT GGC ACC CTA GTC ATT GGA		1145	
Leu Ile Val Ile Trp Glu Pro Ser Pro Ser Gly Thr Leu Val Ile Gly			
295	300	305	
GTC ATC AGT GGA ATT GCT GTT TTT GTC GTC ATC TTG TTC ATT GGA ATT		1193	
Val Ile Ser Gly Ile Ala Val Phe Val Val Ile Leu Phe Ile Gly Ile			
310	315	320	
TTG TTC ATA ATA TTA AGG AAG AGG CAG GGT TCA AGA GGA GCC ATG GGG		1241	
Leu Phe Ile Ile Leu Arg Lys Arg Gln Gly Ser Arg Gly Ala Met Gly			
325	330	335	340
CAC TAC GTC TTA GCT GAA CGT GAG TGACACGCAG CCTGCAGACT CACTGTGGGA		1295	
His Tyr Val Leu Ala Glu Arg Glu			
345			

AGGAGACAAA ACTAGAGACT CAAAGAGGGA GTGCATTTAT GAGCTCTCA TGTTTCAGGA	1355
GAGAGTTGAA CCTAACATA GAAATTGCCT GACGAACCTC TTGATTTAG CCTTCTCTGT	1415
TCATTCCTC AAAAAGATTT CCCCCA	1440

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1440 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 222..1268

- (ix) FEATURE:
 - (A) NAME/KEY: allele
 - (B) LOCATION: replace(1066, "a")
 - (D) OTHER INFORMATION: /phenotype= "Hereditary Hemochromatosis (HH)"
/label= 24d1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GGGGACACTG GATCACCTAG TGTTTCACAA GCAGGTACCT TCTGCTGTAG GAGAGAGAGA	60
ACTAAAGTTC TGAAAGACCT GTTGCTTTTC ACCAGGAAGT TTTACTGGGC ATCTCCTGAG	120
CCTAGGCAAT AGCTGTAGGG TGACTTCTGG AGCCATCCCC GTTTCCCCGC CCCCCAAAAG	180
AAGCGGAGAT TTAACGGGGA CGTGCGGCCA GAGCTGGGGA A ATG GGC CCG CGA Met Gly Pro Arg 1	233
GCC AGG CCG GCG CTT CTC CTC CTG ATG CTT TTG CAG ACC GCG GTC CTG Ala Arg Pro Ala Leu Leu Leu Met Leu Leu Gln Thr Ala Val Leu 5 10 15 20	281
CAG GGG CGC TTG CTG CGT TCA CAC TCT CTG CAC TAC CTC TTC ATG GGT Gln Gly Arg Leu Leu Arg Ser His Ser Leu His Tyr Leu Phe Met Gly 25 30 35	329
GCC TCA GAG CAG GAC CTT GGT CTT TCC TTG TTT GAA GCT TTG GGC TAC Ala Ser Glu Gln Asp Leu Gly Leu Ser Leu Phe Glu Ala Leu Gly Tyr 40 45 50	377
GTG GAT GAC CAG CTG TTC GTG TTC TAT GAT CAT GAG AGT CGC CGT GTG Val Asp Asp Gln Leu Phe Val Phe Tyr Asp His Glu Ser Arg Arg Val 55 60 65	425
GAG CCC CGA ACT CCA TGG GTT TCC AGT AGA ATT TCA AGC CAG ATG TGG Glu Pro Arg Thr Pro Trp Val Ser Ser Arg Ile Ser Ser Gln Met Trp 70 75 80	473

CTG CAG CTG AGT CAG AGT CTG AAA GGG TGG GAT CAC ATG TTC ACT GTT	521
Leu Gln Leu Ser Gln Ser Leu Lys Gly Trp Asp His Met Phe Thr Val	
85 90 95 100	
GAC TTC TGG ACT ATT ATG GAA AAT CAC AAC CAC AGC AAG GAG TCC CAC	569
Asp Phe Trp Thr Ile Met Glu Asn His Asn His Ser Lys Glu Ser His	
105 110 115	
ACC CTG CAG GTC ATC CTG GGC TGT GAA ATG CAA GAA GAC AAC AGT ACC	617
Thr Leu Gln Val Ile Leu Gly Cys Glu Met Gln Glu Asp Asn Ser Thr	
120 125 130	
GAG GGC TAC TGG AAG TAC GGG TAT GAT GGG CAG GAC CAC CTT GAA TTC	665
Glu Gly Tyr Trp Lys Tyr Gly Tyr Asp Gly Gln Asp His Leu Glu Phe	
135 140 145	
TGC CCT GAC ACA CTG GAT TGG AGA GCA GCA GAA CCC AGG GCC TGG CCC	713
Cys Pro Asp Thr Leu Asp Trp Arg Ala Ala Glu Pro Arg Ala Trp Pro	
150 155 160	
ACC AAG CTG GAG TGG GAA AGG CAC AAG ATT CGG GCC AGG CAG AAC AGG	761
Thr Lys Leu Glu Trp Glu Arg His Lys Ile Arg Ala Arg Gln Asn Arg	
165 170 175 180	
GCC TAC CTG GAG AGG GAC TGC CCT GCA CAG CTG CAG CAG TTG CTG GAG	809
Ala Tyr Leu Glu Arg Asp Cys Pro Ala Gln Leu Gln Gln Leu Leu Glu	
185 190 195	
CTG GGG AGA GGT GTT TTG GAC CAA CAA GTG CCT CCT TTG GTG AAG GTG	857
Leu Gly Arg Gly Val Leu Asp Gln Gln Val Pro Pro Leu Val Lys Val	
200 205 210	
ACA CAT CAT GTG ACC TCT TCA GTG ACC ACT CTA CGG TGT CGG GCC TTG	905
Thr His His Val Thr Ser Ser Val Thr Thr Leu Arg Cys Arg Ala Leu	
215 220 225	
AAC TAC TAC CCC CAG AAC ATC ACC ATG AAG TGG CTG AAG GAT AAG CAG	953
Asn Tyr Tyr Pro Gln Asn Ile Thr Met Lys Trp Leu Lys Asp Lys Gln	
230 235 240	
CCA ATG GAT GCC AAG GAG TTC GAA CCT AAA GAC GTA TTG CCC AAT GGG	1001
Pro Met Asp Ala Lys Glu Phe Glu Pro Lys Asp Val Leu Pro Asn Gly	
245 250 255 260	
GAT GGG ACC TAC CAG GGC TGG ATA ACC TTG GCT GTA CCC CCT GGG GAA	1049
Asp Gly Thr Tyr Gln Gly Trp Ile Thr Leu Ala Val Pro Pro Gly Glu	
265 270 275	
GAG CAG AGA TAT ACG TAC CAG GTG GAG CAC CCA GGC CTG GAT CAG CCC	1097
Glu Gln Arg Tyr Thr Tyr Gln Val Glu His Pro Gly Leu Asp Gln Pro	
280 285 290	
CTC ATT GTG ATC TGG GAG CCC TCA CCG TCT GGC ACC CTA GTC ATT GGA	1145
Leu Ile Val Ile Trp Glu Pro Ser Pro Ser Gly Thr Leu Val Ile Gly	
295 300 305	
GTC ATC AGT GGA ATT GCT GTT TTT GTC GTC ATC TTG TTC ATT GGA ATT	1193
Val Ile Ser Gly Ile Ala Val Phe Val Val Ile Leu Phe Ile Gly Ile	
310 315 320	
TTG TTC ATA ATA TTA AGG AAG AGG CAG GGT TCA AGA GGA GCC ATG GGG	1241
Leu Phe Ile Ile Leu Arg Lys Arg Gln Gly Ser Arg Gly Ala Met Gly	

325	330	335	340	
CAC TAC GTC TTA GCT GAA CGT GAG TGACACGCAG CCTGCAGACT CACTGTGGGA				1295
His Tyr Val Leu Ala Glu Arg Glu				
345				
AGGAGACAAA ACTAGAGACT CAAAGAGGGA GTGCATTTAT GAGCTCTTCA TGTTTCAGGA				1355
GAGAGTTGAA CCTAAACATA GAAATTGCCT GACGAACCTCC TTGATTTAG CCTTCTCTGT				1415
TCATTTCTC AAAAAGATTT CCCCA				1440

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1440 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 222..1268

(ix) FEATURE:

- (A) NAME/KEY: allele
- (B) LOCATION: replace(408, "g")
- (D) OTHER INFORMATION: /phenotype= "Hereditary Hemochromatosis (HH)"
/label= 24d2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GGGGACACTG GATCACCTAG TGTTTCACAA GCAGGTACCT TCTGCTGTAG GAGAGAGAGA	60
ACTAAAGITC TGAAAGACCT GTTGCTTTTC ACCAGGAAGT TTTACTGGGC ATCTCCTGAG	120
CCTAGGCAAT AGCTGTAGGG TGACTTCTGG AGCCATCCCC GTTTCCCCGC CCCCCAAAAG	180
AAGCGGAGAT TTAACGGGGA CGTGGGCCA GAGCTGGGGA A ATG GGC CCG CGA	233
Met Gly Pro Arg	
1	
GCC AGG CCG GCG CTT CTC CTC CTG ATG CTT TTG CAG ACC GCG GTC CTG	281
Ala Arg Pro Ala Leu Leu Leu Leu Met Leu Leu Gln Thr Ala Val Leu	
5 10 15 20	
CAG GGG CGC TTG CTG CGT TCA CAC TCT CTG CAC TAC CTC TTC ATG GGT	329
Gln Gly Arg Leu Leu Arg Ser His Ser Leu His Tyr Leu Phe Met Gly	
25 30 35	
GCC TCA GAG CAG GAC CTT GGT CTT TCC TTG TTT GAA GCT TTG GGC TAC	377
Ala Ser Glu Gln Asp Leu Gly Leu Ser Leu Phe Glu Ala Leu Gly Tyr	
40 45 50	
GTG GAT GAC CAG CTG TTC GTG TTC TAT GAT GAG AGT CGC CGT GTG	425

Val	Asp	Asp	Gln	Leu	Phe	Val	Phe	Tyr	Asp	Asp	Glu	Ser	Arg	Arg	Val	
55					60						65					
GAG	CCC	CGA	ACT	CCA	TGG	GTT	TCC	AGT	AGA	ATT	TCA	AGC	CAG	ATG	TGG	473
Glu	Pro	Arg	Thr	Pro	Trp	Val	Ser	Ser	Arg	Ile	Ser	Ser	Gln	Met	Trp	
70					75					80						
CTG	CAG	CTG	AGT	CAG	AGT	CTG	AAA	GGG	TGG	GAT	CAC	ATG	TTC	ACT	GTT	521
Leu	Gln	Leu	Ser	Gln	Ser	Leu	Lys	Gly	Trp	Asp	His	Met	Phe	Thr	Val	
85					90					95					100	
GAC	TTC	TGG	ACT	ATT	ATG	GAA	AAT	CAC	AAC	CAC	AGC	AAG	GAG	TCC	CAC	569
Asp	Phe	Trp	Thr	Ile	Met	Glu	Asn	His	Asn	His	Ser	Lys	Glu	Ser	His	
105					110					115						
ACC	CTG	CAG	GTC	ATC	CTG	GGC	TGT	GAA	ATG	CAA	GAA	GAC	AAC	AGT	ACC	617
Thr	Leu	Gln	Val	Ile	Leu	Gly	Cys	Glu	Met	Gln	Glu	Asp	Asn	Ser	Thr	
120					125					130						
GAG	GGC	TAC	TGG	AAG	TAC	GGG	TAT	GAT	GGG	CAG	GAC	CAC	CTT	GAA	TTC	665
Glu	Gly	Tyr	Trp	Lys	Tyr	Gly	Tyr	Asp	Gly	Gln	Asp	His	Leu	Glu	Phe	
135					140					145						
TGC	CCT	GAC	ACA	CTG	GAT	TGG	AGA	GCA	GCA	GAA	CCC	AGG	GCC	TGG	CCC	713
Cys	Pro	Asp	Thr	Leu	Asp	Trp	Arg	Ala	Ala	Glu	Pro	Arg	Ala	Trp	Pro	
150					155					160						
ACC	AAG	CTG	GAG	TGG	GAA	AGG	CAC	AAG	ATT	CGG	GCC	AGG	CAG	AAC	AGG	761
Thr	Lys	Leu	Glu	Trp	Glu	Arg	His	Lys	Ile	Arg	Ala	Arg	Gln	Asn	Arg	
165					170					175					180	
GCC	TAC	CTG	GAG	AGG	GAC	TGC	CCT	GCA	CAG	CTG	CAG	CAG	TTG	CTG	GAG	809
Ala	Tyr	Leu	Glu	Arg	Asp	Cys	Pro	Ala	Gln	Leu	Gln	Gln	Leu	Leu	Glu	
185					190					195						
CTG	GGG	AGA	GGT	GTT	TTG	GAC	CAA	CAA	GTG	CCT	CCT	TTG	GTG	AAG	GTG	857
Leu	Gly	Arg	Gly	Val	Leu	Asp	Gln	Gln	Val	Pro	Pro	Leu	Val	Lys	Val	
200					205					210						
ACA	CAT	CAT	GTG	ACC	TCT	TCA	GTG	ACC	ACT	CTA	CGG	TGT	CGG	GCC	TTG	905
Thr	His	His	Val	Thr	Ser	Ser	Val	Thr	Thr	Leu	Arg	Cys	Arg	Ala	Leu	
215					220					225						
AAC	TAC	TAC	CCC	CAG	AAC	ATC	ACC	ATG	AAG	TGG	CTG	AAG	GAT	AAG	CAG	953
Asn	Tyr	Tyr	Pro	Gln	Asn	Ile	Thr	Met	Lys	Trp	Leu	Lys	Asp	Lys	Gln	
230					235					240						
CCA	ATG	GAT	GCC	AAG	GAG	TTC	GAA	CCT	AAA	GAC	GTA	TTG	CCC	AAT	GGG	1001
Pro	Met	Asp	Ala	Lys	Glu	Phe	Glu	Pro	Lys	Asp	Val	Leu	Pro	Asn	Gly	
245					250					255					260	
GAT	GGG	ACC	TAC	CAG	GGC	TGG	ATA	ACC	TTG	GCT	GTA	CCC	CCT	GGG	GAA	1049
Asp	Gly	Thr	Tyr	Gln	Gly	Trp	Ile	Thr	Leu	Ala	Val	Pro	Pro	Gly	Glu	
265					270					275						
GAG	CAG	AGA	TAT	ACG	TGC	CAG	GTG	GAG	CAC	CCA	GGC	CTG	GAT	CAG	CCC	1097
Glu	Gln	Arg	Tyr	Thr	Cys	Gln	Val	Glu	His	Pro	Gly	Leu	Asp	Gln	Pro	
280					285					290						
CTC	ATT	GTG	ATC	TGG	GAG	CCC	TCA	CCG	TCT	GGC	ACC	CTA	GTC	ATT	GGA	1145
Leu	Ile	Val	Ile	Trp	Glu	Pro	Ser	Pro	Ser	Gly	Thr	Leu	Val	Ile	Gly	
295					300					305						

GTC ATC AGT GGA ATT GCT GTT TTT GTC GTC ATC TTG TTC ATT GGA ATT	1193
Val Ile Ser Gly Ile Ala Val Phe Val Val Ile Leu Phe Ile Gly Ile	
310 315 320	
TTG TTC ATA ATA TTA AGG AAG AGG CAG GGT TCA AGA GGA GCC ATG GGG	1241
Leu Phe Ile Ile Leu Arg Lys Arg Gln Gly Ser Arg Gly Ala Met Gly	
325 330 335 340	
CAC TAC GTC TTA GCT GAA CGT GAG TGACACGCAG CCTGCAGACT CACTGTGGGA	1295
His Tyr Val Leu Ala Glu Arg Glu	
345	
AGGAGACAAA ACTAGAGACT CAAAGAGGGA GTGCATTTAT GAGCTCTTCA TGTTTCAGGA	1355
GAGAGTTGAA CCTAACATA GAAATTGCCT GACGAACCTCC TTGATTTAG CCTTCTCTGT	1415
TCATTCCTC AAAAGATTT CCCCA	1440

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1440 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 222..1268
- (ix) FEATURE:
 - (A) NAME/KEY: allele
 - (B) LOCATION: replace(408, "g")
 - (D) OTHER INFORMATION: /phenotype= "Hereditary Hemochromatosis (HH)"
/label= 24d2
- (ix) FEATURE:
 - (A) NAME/KEY: allele
 - (B) LOCATION: replace(1066, "a")
 - (D) OTHER INFORMATION: /phenotype= "Hereditary Hemochromatosis (HH)"
/label= 24d1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GGGGACACTG GATCACCTAG TGTTTCACAA GCAGGTACCT TCTGCTGTAG GAGAGAGAGA	60
ACTAAAGTTTC TGAAAGACCT GTTGCTTTTC ACCAGGAAGT TTTACTGGGC ATCTCCTGAG	120
CCTAGGCAAT AGCTGTAGGG TGACTTCTGG AGCCATCCCC GTTTCCCCGC CCCCCAAAAG	180
AAGCGGAGAT TTAACGGGGA CGTGCAGGCCA GAGCTGGGGA A ATG GGC CCG CGA	233
Met Gly Pro Arg	

GCC AGG CCG GCG CTT CTC CTC CTG ATG CTT TTG CAG ACC GCG GTC CTG Ala Arg Pro Ala Leu Leu Leu Met Leu Leu Gln Thr Ala Val Leu 5 10 15 20	281
CAG GGG CGC TTG CTG CGT TCA CAC TCT CTG CAC TAC CTC TTC ATG GGT Gln Gly Arg Leu Leu Arg Ser His Ser Leu His Tyr Leu Phe Met Gly 25 30 35	329
GCC TCA GAG CAG GAC CTT GGT CTT TCC TTG TTT GAA GCT TTG GGC TAC Ala Ser Glu Gln Asp Leu Gly Leu Ser Leu Phe Glu Ala Leu Gly Tyr 40 45 50	377
GTG GAT GAC CAG CTG TTC GTG TTC TAT GAT GAT GAG AGT CGC CGT GTG Val Asp Asp Gln Leu Phe Val Phe Tyr Asp Asp Glu Ser Arg Arg Val 55 60 65	425
GAG CCC CGA ACT CCA TGG GTT TCC AGT AGA ATT TCA AGC CAG ATG TGG Glu Pro Arg Thr Pro Trp Val Ser Ser Arg Ile Ser Ser Gln Met Trp 70 75 80	473
CTG CAG CTG AGT CAG AGT CTG AAA GGG TGG GAT CAC ATG TTC ACT GTT Leu Gln Leu Ser Gln Ser Leu Lys Gly Trp Asp His Met Phe Thr Val 85 90 95 100	521
GAC TTC TGG ACT ATT ATG GAA AAT CAC AAC CAC AGC AAG GAG TCC CAC Asp Phe Trp Thr Ile Met Glu Asn His Asn His Ser Lys Glu Ser His 105 110 115	569
ACC CTG CAG GTC ATC CTG GGC TGT GAA ATG CAA GAA GAC AAC AGT ACC Thr Leu Gln Val Ile Leu Gly Cys Glu Met Gln Glu Asp Asn Ser Thr 120 125 130	617
GAG GGC TAC TGG AAG TAC GGG TAT GAT GGG CAG GAC CAC CTT GAA TTC Glu Gly Tyr Trp Lys Tyr Gly Tyr Asp Gly Gln Asp His Leu Glu Phe 135 140 145	665
TGC CCT GAC ACA CTG GAT TGG AGA GCA GCA GAA CCC AGG GCC TGG CCC Cys Pro Asp Thr Leu Asp Trp Arg Ala Ala Glu Pro Arg Ala Trp Pro 150 155 160	713
ACC AAG CTG GAG TGG GAA AGG CAC AAG ATT CGG GCC AGG CAG AAC AGG Thr Lys Leu Glu Trp Glu Arg His Lys Ile Arg Ala Arg Gln Asn Arg 165 170 175 180	761
GCC TAC CTG GAG AGG GAC TGC CCT GCA CAG CTG CAG CAG TTG CTG GAG Ala Tyr Leu Glu Arg Asp Cys Pro Ala Gln Leu Gln Leu Leu Glu 185 190 195	809
CTG GGG AGA GGT GTT TTG GAC CAA CAA GTG CCT CCT TTG GTG AAG GTG Leu Gly Arg Gly Val Leu Asp Gln Gln Val Pro Pro Leu Val Lys Val 200 205 210	857
ACA CAT CAT GTG ACC TCT TCA GTG ACC ACT CTA CGG TGT CGG GCC TTG Thr His His Val Thr Ser Ser Val Thr Thr Leu Arg Cys Arg Ala Leu 215 220 225	905
AAC TAC TAC CCC CAG AAC ATC ACC ATG AAG TGG CTG AAG GAT AAG CAG Asn Tyr Tyr Pro Gln Asn Ile Thr Met Lys Trp Leu Lys Asp Lys Gln 230 235 240	953
CCA ATG GAT GCC AAG GAG TTC GAA CCT AAA GAC GTA TTG CCC AAT GGG Pro Met Asp Ala Lys Glu Phe Glu Pro Lys Asp Val Leu Pro Asn Gly	1001

245	250	255	260	
GAT GGG ACC TAC CAG GGC TGG ATA ACC TTG GCT GTA CCC CCT GGG GAA				1049
Asp Gly Thr Tyr Gln Gly Trp Ile Thr Leu Ala Val Pro Pro Gly Glu				
265	270	275		
GAG CAG AGA TAT ACG TAC CAG GTG GAG CAC CCA GGC CTG GAT CAG CCC				1097
Glu Gln Arg Tyr Thr Tyr Gln Val Glu His Pro Gly Leu Asp Gln Pro				
280	285	290		
CTC ATT GTG ATC TGG GAG CCC TCA CCG TCT GGC ACC CTA GTC ATT GGA				1145
Leu Ile Val Ile Trp Glu Pro Ser Pro Ser Gly Thr Leu Val Ile Gly				
295	300	305		
GTC ATC AGT GGA ATT GCT GTT TTT GTC GTC ATC TTG TTC ATT GGA ATT				1193
Val Ile Ser Gly Ile Ala Val Phe Val Val Ile Leu Phe Ile Gly Ile				
310	315	320		
TTG TTC ATA ATA TTA AGG AAG AGG CAG GGT TCA AGA GGA GCC ATG GGG				1241
Leu Phe Ile Ile Leu Arg Lys Arg Gln Gly Ser Arg Gly Ala Met Gly				
325	330	335	340	
CAC TAC GTC TTA GCT GAA CGT GAG TGACACGCAG CCTGCAGACT CACTGTGGGA				1295
His Tyr Val Leu Ala Glu Arg Glu				
345				
AGGAGACAAA ACTAGAGACT CAAAGAGGGA GTGCATTTAT GAGCTCTTCA TGTTTCAGGA				1355
GAGAGTTGAA CCTAACATA GAAATTGCCT GACGAACCTCC TTGATTTAG CCTTCTCTGT				1415
TCATTTCTC AAAAGATTT CCCCA				1440

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TGGCAAGGGT AACAGATCC

20

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CTCAGGCACT CCTCTCAACC

20

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /mod_base= OTHER
/note= "N = 5'-biotinylated guanine
(bio-G)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

NGAAGAGCAG AGATATACGT G

21

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /mod_base= OTHER
/note= "N = 5'-biotinylated guanine
(bio-G)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

NGAAGAGCAG AGATATACGT A

21

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:
 (A) NAME/KEY: modified_base
 (B) LOCATION: 1
 (D) OTHER INFORMATION: /mod_base= OTHER
 /note= "N = 5'-phosphorylated cytosine
 (p-C)"

(ix) FEATURE:
 (A) NAME/KEY: modified_base
 (B) LOCATION: 18
 (D) OTHER INFORMATION: /mod_base= OTHER
 /note= "N = 3'-digoxigenin-conjugated
 guanine (G-dig)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

NCAGGTGGAG CACCCAGN

18

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CTGAAAGGGT GGGATCACAT

20

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CAAGGAGTTC GTCAGGCAAT

20

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 517 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 1..517
- (D) OTHER INFORMATION: /note= "normal or wild-type (unaffected) genomic sequence surrounding variant for 24d1(G) allele corresponding to positions 5507-6023 of genomic sequence containing the HH gene (SEQ ID NO:1)"

(ix) FEATURE:

- (A) NAME/KEY: allele
- (B) LOCATION: replace(328, "g")
- (D) OTHER INFORMATION: /phenotype= "normal or wild-type (unaffected)"
/label= 24d1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TATTTCTTC	CTCCAACCTA	TAGAAGGAAG	TGAAAGTTCC	AGTCTTCCTG	GCAAGGGTAA	60
ACAGATCCCC	TCTCCTCATC	CTTCCTCTTT	CCTGTCAAGT	GCCTCCTTTG	GTGAAGGTGA	120
CACATCATGT	GACCTCTTCA	GTGACCACTC	TACGGTGTGCG	GGCCTTGAAC	TACTACCCCC	180
AGAACATCAC	CATGAAGTGG	CTGAAGGATA	AGCAGCCAAT	GGATGCCAAG	GAGTCGAAC	240
CTAAAGACGT	ATTGCCAAT	GGGGATGGGA	CCTACCAGGG	CTGGATAACC	TTGGCTGTAC	300
CCCCTGGGGA	AGAGCAGAGA	TATACGTGCC	AGGTGGAGCA	CCCAGGCCTG	GATCAGCCCC	360
TCATTGTGAT	CTGGGGTATG	TGACTGATGA	GAGCCAGGAG	CTGAGAAAAT	CTATTGGGGG	420
TTGAGAGGAG	TGCCTGAGGA	GGTAATTATG	GCAGTGAGAT	GAGGATCTGC	TCTTGTTAG	480
GGGGTGGGCT	GAGGGTGGCA	ATCAAAGGCT	TTAACTT			517

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 517 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 1..517
- (D) OTHER INFORMATION: /note= "genomic sequence surrounding variant for 24d1(A) allele corresponding to positions 5507-6023 of genomic sequence containing the HH gene (SEQ ID NO:3)"

(ix) FEATURE:

- (A) NAME/KEY: allele
- (B) LOCATION: replace(328, "a")
- (D) OTHER INFORMATION: /phenotype= "Hereditary Hemochromatosis"

(HH) "
/label= 24d1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

TATTTCCCTTC	CTCCAACCTA	TAGAAGGAAG	TGAAAGTTCC	AGTCTTCCTG	GCAAGGGTAA	60
ACAGATCCCC	TCTCCTCATC	CTTCCTCTTT	CCTGTCAAGT	GCCTCCTTTG	GTGAAGGGTGA	120
CACATCATGT	GACCTCTTCA	GTGACCACTC	TACGGTGTG	GGCCTTGAAC	TACTACCCCC	180
AGAACATCAC	CATGAAGTGG	CTGAAGGATA	AGCAGCCAAT	GGATGCCAAG	GAGTCGAAC	240
CTAAAGACGT	ATTGCCAAT	GGGGATGGGA	CCTACCAGGG	CTGGATAACC	TTGGCTGTAC	300
CCCCTGGGGA	AGAGCAGAGA	TATACGTACC	AGGTGGAGCA	CCCAGGCCTG	GATCAGCCCC	360
TCATTGTGAT	CTGGGGTATG	TGACTGATGA	GAGCCAGGAG	CTGAGAAAAT	CTATTGGGGG	420
TTGAGAGGAG	TGCCTGAGGA	GGTAATTATG	GCAGTGAGAT	GAGGATCTGC	TCTTGTTAG	480
GGGGTGGGCT	GAGGGTGGCA	ATCAAAGGCT	TTAACTT			517

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 361 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ix) FEATURE:

- (A) NAME/KEY: Protein
- (B) LOCATION: 1..361
- (D) OTHER INFORMATION: /note= "Rabbit leukocyte antigen (RLA)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Met	Gly	Ser	Ile	Pro	Pro	Arg	Thr	Leu	Leu	Leu	Leu	Leu	Ala	Gly	Ala
1				5				10						15	
Leu	Thr	Leu	Lys	Asp	Thr	Gln	Ala	Gly	Ser	His	Ser	Met	Arg	Tyr	Phe
				20			25					30			
Tyr	Thr	Ser	Val	Ser	Arg	Pro	Gly	Leu	Gly	Glu	Pro	Arg	Phe	Ile	Ile
	35				40					45					
Val	Gly	Tyr	Val	Asp	Asp	Thr	Gln	Phe	Val	Arg	Phe	Asp	Ser	Asp	Ala
	50				55			60							
Ala	Ser	Pro	Arg	Met	Glu	Gln	Arg	Ala	Pro	Trp	Met	Gly	Gln	Val	Glu
	65			70				75		80					
Pro	Glu	Tyr	Trp	Asp	Gln	Gln	Thr	Gln	Ile	Ala	Lys	Asp	Thr	Ala	Gln
	85				90					95					

Thr Phe Arg Val Asn Leu Asn Thr Ala Leu Arg Tyr Tyr Asn Gln Ser
 100 105 110

Ala Ala Gly Ser His Thr Phe Gln Thr Met Phe Gly Cys Glu Val Trp
 115 120 125

Ala Asp Gly Arg Phe Phe His Gly Tyr Arg Gln Tyr Ala Tyr Asp Gly
 130 135 140

Ala Asp Tyr Ile Ala Leu Asn Glu Asp Leu Arg Ser Trp Thr Ala Ala
 145 150 155 160

Asp Thr Ala Ala Gln Asn Thr Gln Arg Lys Trp Glu Ala Ala Gly Glu
 165 170 175

Ala Glu Arg His Arg Ala Tyr Leu Glu Arg Glu Cys Val Glu Trp Leu
 180 185 190

Arg Arg Tyr Leu Glu Met Gly Lys Glu Thr Leu Gln Arg Ala Asp Pro
 195 200 205

Pro Lys Ala His Val Thr His His Pro Ala Ser Asp Arg Glu Ala Thr
 210 215 220

Leu Arg Cys Trp Ala Leu Gly Phe Tyr Pro Ala Glu Ile Ser Leu Thr
 225 230 235 240

Trp Gln Arg Asp Gly Glu Asp Gln Thr Gln Asp Thr Glu Leu Val Glu
 245 250 255

Thr Arg Pro Gly Gly Asp Gly Thr Phe Gln Lys Trp Ala Ala Val Val
 260 265 270

Val Pro Ser Gly Glu Glu Gln Arg Tyr Thr Cys Arg Val Gln His Glu
 275 280 285

Gly Leu Pro Glu Pro Leu Thr Leu Thr Trp Glu Pro Pro Ala Gln Pro
 290 295 300

Thr Ala Leu Ile Val Gly Ile Val Ala Gly Val Leu Gly Val Leu Leu
 305 310 315 320

Ile Leu Gly Ala Val Val Ala Val Val Arg Arg Lys Lys His Ser Ser
 325 330 335

Asp Gly Lys Gly Gly Arg Tyr Thr Pro Ala Ala Gly Gly His Arg Asp
 340 345 350

Gln Gly Ser Asp Asp Ser Leu Met Pro
 355 360

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 365 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

(ix) FEATURE:

- (A) NAME/KEY: Protein
- (B) LOCATION: 1..365
- (D) OTHER INFORMATION: /note= "Human Major Histocompatibility Class I (MHC) protein"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Met	Ala	Val	Met	Ala	Pro	Arg	Thr	Leu	Val	Leu	Leu	Leu	Ser	Gly	Ala
1				5					10				15		
Leu	Ala	Leu	Thr	Gln	Thr	Trp	Ala	Gly	Ser	His	Ser	Met	Arg	Tyr	Phe
		20					25					30			
Phe	Thr	Ser	Val	Ser	Arg	Pro	Gly	Arg	Gly	Glu	Pro	Arg	Phe	Ile	Ala
	35				40				45						
Val	Gly	Tyr	Val	Asp	Asp	Thr	Gln	Phe	Val	Arg	Phe	Asp	Ser	Asp	Ala
	50				55				60						
Ala	Ser	Gln	Arg	Met	Glu	Pro	Arg	Ala	Pro	Trp	Ile	Glu	Gln	Glu	Gly
	65				70				75			80			
Pro	Glu	Tyr	Trp	Asp	Gly	Glu	Thr	Arg	Lys	Val	Lys	Ala	His	Ser	Gln
		85				90						95			
Thr	His	Arg	Val	Asp	Leu	Gly	Thr	Leu	Arg	Gly	Tyr	Tyr	Asn	Gln	Ser
	100						105					110			
Glu	Ala	Gly	Ser	His	Thr	Leu	Gln	Met	Met	Phe	Gly	Cys	Asp	Val	Gly
	115					120					125				
Ser	Asp	Trp	Arg	Phe	Leu	Arg	Gly	Tyr	His	Gln	Tyr	Ala	Tyr	Asp	Gly
	130					135				140					
Lys	Asp	Tyr	Ile	Ala	Leu	Lys	Glu	Asp	Leu	Arg	Ser	Trp	Thr	Ala	Ala
	145					150				155			160		
Asp	Met	Ala	Ala	Gln	Thr	Thr	Lys	His	Lys	Trp	Glu	Ala	Ala	His	Val
					165				170			175			
Ala	Glu	Gln	Leu	Arg	Ala	Tyr	Leu	Glu	Gly	Thr	Cys	Val	Glu	Trp	Leu
			180				185					190			
Arg	Arg	Tyr	Leu	Glu	Asn	Gly	Lys	Glu	Thr	Leu	Gln	Arg	Thr	Asp	Ala
	195					200					205				
Pro	Lys	Thr	His	Met	Thr	His	His	Ala	Val	Ser	Asp	His	Glu	Ala	Thr
	210					215					220				
Leu	Arg	Cys	Trp	Ala	Leu	Ser	Phe	Tyr	Pro	Ala	Glu	Ile	Thr	Leu	Thr
	225					230				235			240		
Trp	Gln	Arg	Asp	Gly	Glu	Asp	Gln	Thr	Gln	Asp	Thr	Glu	Leu	Val	Glu
		245				250						255			
Thr	Arg	Pro	Ala	Gly	Asp	Gly	Thr	Phe	Gln	Lys	Trp	Ala	Ala	Val	Val
	260						265					270			
Val	Pro	Ser	Gly	Gln	Glu	Gln	Arg	Tyr	Thr	Cys	His	Val	Gln	His	Glu
			275			280					285				

Gly Leu Pro Lys Pro Leu Thr Leu Arg Trp Glu Pro Ser Ser Gln Pro
 290 295 300

Thr Ile Pro Ile Val Gly Ile Ile Ala Gly Leu Val Leu Phe Gly Ala
 305 310 315 320

Val Ile Thr Gly Ala Val Val Ala Ala Val Met Trp Arg Arg Lys Ser
 325 330 335

Ser Asp Arg Lys Gly Gly Ser Tyr Ser Gln Ala Ala Ser Ser Asp Ser
 340 345 350

Ala Gln Gly Ser Asp Val Ser Leu Thr Ala Cys Lys Val
 355 360 365

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

ACATGGTTAA GGCCTGTTGC

20

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GCCACATCTG GCTTGAAATT

20

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

(A) NAME/KEY: modified_base
 (B) LOCATION: 1

(D) OTHER INFORMATION: /mod_base= OTHER
 /note= "N = 5'-biotinylated adenine
 (bio-A)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

NGCTGTTCGT GTTCTATGAT C

21

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:
 (A) NAME/KEY: modified_base
 (B) LOCATION: 1
 (D) OTHER INFORMATION: /mod_base= OTHER
 /note= "N = 5'-biotinylated adenine
 (bio-A)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

NGCTGTTCGT GTTCTATGAT G

21

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 19 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:
 (A) NAME/KEY: modified_base
 (B) LOCATION: 1
 (D) OTHER INFORMATION: /mod_base= OTHER
 /note= "N = 5'-phosphorylated adenine
 (p-A)"

(ix) FEATURE:
 (A) NAME/KEY: modified_base
 (B) LOCATION: 19
 (D) OTHER INFORMATION: /mod_base= OTHER
 /note= "N = 3'-digoxigenin-conjugated
 adenine (A-dig)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

NTGAGAGTCG CCGTGTGGN

19

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GGAAGAGCAG AGATATACTGT GCCAGGTGGA GCACCCAGG

39

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GGAAGAGCAG AGATATACTGT ACCAGGTGGA GCACCCAGG

39

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

CAAAAGAAGC GGAGATTAA CG

22

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

AGATTTAACG GGGACGTGC

19

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

AGAGGTCACA TGATGTGTCA CC

22

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

AGGAGGCACT TGTTGGTCC

19

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

AAAATCACAA CCACAGCAAA G

21

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

TTCCCCACAGT GAGTCTGCAG

20

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CAATGGGGAT GGGACCTAC

19

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

ATATACGTGC CAGGTGGAGC

20

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

CCTCTTCACA ACCCCTTTCA

20

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

CATAGCTGTG CAACTCACAT CA

22

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

AGCTGTTCGT GTTCTATGAT CATGAGAGTC GCCGTGTGGA

40

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

AGCTGTTCGT GTTCTATGAT GATGAGAGTC GCCGTGTGGA

40

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

TGTTCTATGA TCATGAGAGT CGCCGTGTGG AG

32

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

TGTTCTATGA TCATGAGTGT CGCCGTGTGG AG

32